LOCOMOTION

Python translation and e-Handbook for model sharing and transparency

WP 11, Task 11.2, D.11.2

August 2023

LOW-CARBON SOCIETY: AN ENHANCED MODELLING TOOL FOR THE TRANSITION TO SUSTAINABILITY (LOCOMOTION)

H2020-LC-CLA-2018-2

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821105.

DOCUMENT HISTORY

Project Acronym	LOCOMOTION	
Project title	Low-carbon society: an enhanced modelling tool for the transition to sustainability	
Project coordination	Universidad de Valladolid (Spain)	
Project duration	1st June 2019 – 30th November 2023	
Deliverable No.	D11.2. Python translation and e-Handbook for model sharing and transparency	
Dissemination level	Confidential (CO) / Public (PU)	
Status		In progress
		To be verified by other WPs
	*	Final
Due date of deliverable	31st of August 2023	
Delivery date	31st of August 2023	
Version	v.0.13	
Work package	WP11 - User engagement: Decision-making, awareness tools and gamification	
Lead beneficiary	CARTIF	
Author (s)	Roger Samsó (CREAF), Iván Ramos (CARTIF), Noelia Ferreras (CARTIF), Mohamed Lifi (UVa), Margarita Mediavilla (UVa), Luis LLases (UVa), Ignacio de Blas Sanz (UVa), Iñigo Capellán-Pérez (UVa), Gonzalo Parrado-Hernado (UVa), Marta Calleja (UVa), David Alvarez Antelo (UVa), Natasa Markovska (SDEWES), Iñaki Arto (BC3), Tomás Calheiros (FC.ID), Stavroula Papagianni (CRES), Ole van Alen (INN), Lukas Eggler (AEA)	

Date	Ver.	Contributors	Comment
12/06/2023	0.1	Roger Samsó (CREAF)	Table of contents
12/06/2023	0.2	Roger Samsó (CREAF)	Writing the Executive Summary, Introduction, Translating the WILIAM model to Python, The pywiliam interface and the Annex sections.
22/06/2021	0.3	Roger Samsó (CREAF)	Added index of the e-handbook section and included the Land and Water module content.
01/05/2023	0.4	Mohamed Lifi (UVa) & Margarita Mediavilla (UVa)	Land and Water module chapter for e-handbook
01/07/2023	0.5	Natasa Markovska (SDEWES)	Scenarios section for e-handbook
05/07/2023	0.6	Luís Llases (BS3)	Financial module chapter for e-handbook

06/07/2023	0.6	Iván Ramos (CARTIF)	Improvements in the Financial and environmental module for e-handbook
10/07/2023	0.7	Iñaki Arto (BC3) & Noelia Ferreras (CARTIF)	Economy module and Climate module chapters for e-handbook respectively
10/07/2023	0.8	Gonzalo Parrado-Hernando (UVa)	Demography module
12/07/2023	0.9	Roger Samsó (CREAF)	Accept all changes and general revision. Add content for the Chapter on Versions and Interfaces of the WILIAM model, in the e-handbook section.
13/07/2023	0.10	Mohamed Lifi (UVa), Ignacio de Blas Sanz (UVa), and Iñigo Capellán-Pérez (UVa)	Added "Integrated assessment models" & "WILIAM model" sections.
14/07/2023	0.11	Gonzalo Parrado-Hernando (UVa)	Society module
19/07/2023 & 26/07/2023	0.12	Stavroula Papagianni (CRES)	Model transparency & corrections
16/08/2023	0.12	Iñigo Capellán (UVa), Juan Manuel Campos (UVa)	Improvements in the Energy module
29/08/2023	0.12	Ole van Alen (INN)	Materials module
29/08/2023	0.13	Iván Ramos (CARTIF), Stavroula Papagianni (CRES)	Modules integration and deliverable alignment. References

COPYRIGHT

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS AND ACRONYMS

Acronym	Description
AIM IMAGE	Integrated Model to Assess the Global Environment
AMOC	Atlantic Meridional Overturning Circulation
AMR	Abstract Model Representation
BECCS	BioEnergy Carbon Capture and Storage
CGE	Computable General Equilibrium models
CLI	Command Line Interface
DICE	Dynamic Integrated Climate-Economy Model
EROI	Energy Return on Investment
FUND	Climate Framework for Uncertainty, Negotiation and Distribution
GDPpc	Gross Domestic Product per capita
GFCF	Gross Fixed Capital Formation
GHG	Greenhouse gases
GUI	Graphical User Interface
GWP	Global Warming Potential
IAM	Integrated Assessment Model
IAMC	Integrated Assessment Model Consortium
IO	Input-Output
IPCC	International Panel on Climate Change
LULUCF	Land Use, Land-Use Change and Forestry
ME	Macro-Econometric models
MERGE	Integrated Assessment Model for Global Climate Change
MESSAGE	Model for Energy Supply Strategy Alternatives and their General Environmental Impact
MRIO	Multi-Regional Input-Output
O&M	Operation and Maintenance
PIH	Permanent Income Hypothesis
REMIND	Regional Model of Investments and Development
SSP	Shared Socioeconomic Pathways
WITCH	World Induced Technical Change Hybrid Model
RURR	Remaining resources and reserves of materials

EXECUTIVE SUMMARY

In this deliverable we present all the work carried out during the development of Task 11.2 entitled *Model opening-up and transparency strategy*. Throughout this document we present the work carried out to translate the Vensim® version of the WILIAM model to Python code and the WILIAM e-handbook.

Following the steps initiated during the MEDEAS H2020 project, the WILIAM model is being prototyped in the proprietary Vensim® software. However, to improve transparency and to provide an ease access to the model, the objective of the Task 11.2 whose results are included in this deliverable is to produce a Python open-source version of the WILIAM model. Similar to what was done in MEDEAS project, in LOCOMOTION we also use the PySD library to make the translation to Python. However, in MEDEAS we developed ad-hoc code, to compensate for the missing features of PySD at that time. The newly developed code was specific to the MEDEAS model and could not be used for WILIAM or for any other model built using Vensim®. Besides, the translated model had several issues that made it slow, less clear and more difficult to maintain. Therefore, in this task we addressed all these issues directly in the library, which has resulted in 40 PySD releases since December 2020, being the version 3.10.0, the current release at the time of writing this deliverable.

The current version of PySD allows translating the WILIAM model in an efficient way, and produces clean, maintainable and efficient Python code. The *pywiliam* model is available in a public repository and is released together with a layer of additional code developed to facilitate the parametrisation of the model, launching simulations and processing and plotting the simulation results.

Finally, this Deliverable includes the WILIAM e-handbook, which will serve as a simplified documentation of the approach, conceptual models and hypothesis behind the WILIAM model. WILIAM e-handbook contains the most relevant information of WILIAM model and its modules, instructions or advice on how to use the different model versions and information on the other developments of LOCOMOTION as the associated tools and Data Client.

1. INTRODUCTION

The objectives of Task 11.2 are 1) to release a Python version of the WILIAM model under an open-source license; and 2) to write an e-handbook describing the main features and hypothesis behind the WILIAM model and all tools around it.

WILIAM is a large Integrated Assessment Model (IAM), which is being developed in the proprietary Vensim® software by several modelling teams in the LOCOMOTION consortium. Several intermediate WILIAM versions have been internally released since development started. According to the previous experience, manually translating the model to Python was seen as an unviable since development started.

PySD is an open-source Python library that facilitates the translation of models written in Vensim® and Stella to Python. It also includes an execution backend, to allow simulating the translated models. This library was already used during the H2020 MEDEAS project (2016-2019) to translate the MEDEAS model to Python. However, the version of PySD used back then (v0.10) was far from being production ready. Indeed, a pre-processing of the Vensim® version of the MEDEAS model was required before running PySD, and a postprocessing of the resulting Python code was required for it to properly run. Even with those additional steps, the resulting Python model was difficult to read and maintain, and some differences were present between the simulation results obtained in Vensim® and Python. In terms of performance, the Python model run significantly slower than the original Vensim® model (45 minutes vs ~1 minute), which also discouraged potential users.

Likewise, the functions that could be used when modelling in Vensim® was limited to those already supported by PySD. This was problematic, as it slowed down the development of the Vensim® model, requiring the use of more or less convenient workarounds.

According to the previous, rather than creating ad-hoc solutions for the translation of the WILIAM model based on PySD (as was done in MEDEAS), in this work we put all efforts into improving PySD itself. The result is a new version of PySD (v3.10 at the time of writing), which not only can translate the WILIAM model and run it efficiently but can also be used to translate any other System Dynamics (SD) model. Therefore, the work done in this Task will be beneficial for the entire SD community and resulted in a publication in the Journal of Open-Source Software, coinciding with release 3.0 of PySD.

In addition to all stated improvements to the PySD library, in this task we also created specific tools to facilitate the parametrisation and execution of the *pywiliam* model, and the visualisation and processing of the results. In this document we refer to this piece of code as the *pywiliam interface*, which is released together with the *pywiliam* model, under the MIT open-source license and in the same repository.

Finally, Task 11.2 coordinated the writing of the e-handbook, which will be used for dissemination of the model during demonstration events and modelling workshops. It includes a description of the modelling approaches and main hypotheses used in each individual module, and for the whole WILIAM model.

To cover all the developments previously exposed, this document is organised as follows: Section 2 is dedicated to present all the improvements made in PySD; Section 3 describes the code developed as the interface to the *pymedeas* model and the plotting tool; and Section 4 shows the e-handbook.

2. TRANSLATING THE WILIAM MODEL TO PYTHON

The first goal of Task 11.2 was to translate the WILIAM model, originally developed using the proprietary Vensim® software, to Python. This was motivated by the need to increase transparency of the modelling approach and to increase accessibility. An additional advantage of having the model translated to Python is that the user has access to the full Python ecosystem, which offers powerful open-source tools for data visualization, sensitivity analysis, graph theory, machine learning and other data analysis tasks, which are unavailable in Vensim® (Martin-Martinez et al., 2022). Also, Vensim® is only available for Windows and Mac, while Python may be installed in all platforms.

As mentioned in the Introduction (Section 1), we used the open-source PySD library (Martin-Martinez et al., 2022) to automatically translate the WILIAM model from Vensim® to Python, as was already done in the MEDEAS project. The use of PySD is justified not only by the experience gained in MEDEAS with this tool, but also by the fact that the development of the WILIAM model was expected to last until the last stages of the project, which would have left very little time to work on a manual translation. Therefore, using a tool that automatically translates Vensim® code to Python, was seen as the most effective approach.

In MEDEAS, version 0.10.0 of PySD was used, which had severe limitations in terms of missing features, bugs and execution performance. Therefore, at that time the focus was put on overcoming such limitations by developing ad-hoc code to preprocess the Vensim® code before translating it with PySD, and then postprocess the resulting Python code to make it runnable. Although many bugs found in PySD were reported along the way, this approach was very time consuming and brought little added value to the PySD project and to the SD community. According to the previous, in LOCOMOTION we opted to work on improving PySD itself, rather than developing ad-hoc tools to counteract its initial limitations.

Development started in September 2020, departing from version 0.10.0 of PySD. While the WILIAM model was still in a very early stage of development, and since it was to be based on the MEDEAS model, we started by identifying which new features would need to be added in PySD to translate the MEDEAS model. Then, from the first internal release of WILIAM we started using WILIAM to test the PySD library. Until September 2022, the focus was on fixing bugs in PySD and adding support for all Vensim® functions used in WILIAM. From September 2022 onwards, the focus switched to reporting issues and bugs in the Vensim® version of the WILIAM model which prevented the translation using PySD.

From version 0.10.0 to the current version 3.10.0, the improvements made in the PySD library may be classified in the following topics:

1. Increasing the number of supported Vensim® functions
2. Rewriting and restructuring the library
3. Adding new features and options
4. Improving resulting code clarity to facilitate maintenance
5. Improving user feedback and error messages
6. Increasing test coverage
7. Fixing bugs
8. Increasing execution performance
9. Improving documentation
10. Other relevant changes

For the sake of brevity, only the most relevant improvements on each of those categories are described in Section 2.1: Improvements to the PySD library. Note that for practical reasons those improvements will not be reported in strict chronological order, nor by corresponding release number. If the reader wants to have this additional information, they may refer to the Annex for PySD releases between 0.11.0 and 2.2.4 and to the *What's new* section of the documentation (https://pysd.readthedocs.io/en/master/whats_new.html) for subsequent releases.

Additionally, it is worth mentioning that the PySD library went through an open peer review process during the publication of the paper in which we described the new structure of the library from release 3.0.0 in the Journal of Open Source Software (Martin-Martinez et al., 2022). The whole review process, including the reviewers' requests and our responses may be consulted in this link: https://github.com/openjournals/joss-reviews/issues/4329. This review helped us improve the quality of the whole library, and most notably the quality, clarity and completeness of the documentation.

On the other hand, the procedure followed along the whole WILIAM development cycle, which ended up with a version of the model that could be directly translated to Python using PySD, is described in Section 2.2:Translation of the WILIAM model.

2.1. IMPROVEMENTS TO THE PYSD LIBRARY

2.1.1. INCREASING THE NUMBER OF SUPPORTED VENSIM® FUNCTIONS

The current list of supported Vensim® functions can be found in this section of the documentation (https://pysd.readthedocs.io/en/master/structure/vensim_translation.html?highlight=supported%20Vensim#functions).

Since PySD 0.10.0, the support for the following Vensim® functions was added:

- SAMPLE IF TRUE
- INVERT MATRIX
- FORECAST
- :NA:
- ELMCOUNT
- :EXCEPT:
- VECTOR SORT ORDER
- VECTOR RANK
- VECTOR REORDER
- VECTOR SELECT
- GET TIME VALUE
- ALLOCATE BY PRIORITY
- ALLOCATE AVAILABLE
- POWER
- TABBED ARRAYS

2.1.2. REWRITING AND RESTRUCTURING THE LIBRARY

PySD was originally created in 2014 as a side project of James Houghton (Houghton & Siegel, 2015), who at that time was a PhD student at the Massachusetts Institute of Technology. From then, there have been several contributions from the SD community, which have greatly improved the features of the library. However, after working on it for a few months, adding new features was becoming increasingly complex at every new release. Code maintenance was becoming equally complex.

In the process of translating a Vensim® model to Python using PySD, three main steps are involved: 1) reading the model file (.mdl) and parsing each expression, 2) creating strings of Python code that match the parsed expressions, and 3) writing the strings into a Python file (with *.py* extension).

In PySD versions prior to 3.0.0, steps 1 and 2 were carried out in parallel, hence Python strings were already created for certain expressions before the whole code had been parsed. This was problematic, as a single expression in Vensim® may be written in multiple lines (see the *my value* expression in Figure 1).

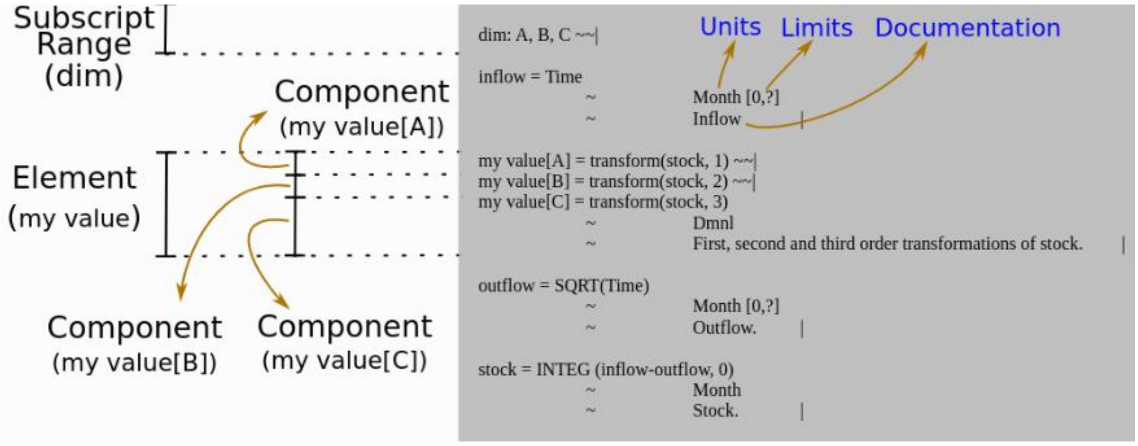


Figure 1: Screen capture of the code of a model written in Vensim®. Adapted from: https://pysd.readthedocs.io/en/master/structure/vensim_translation.html?highlight=supported%20Vensim#translation-workflow

Moreover, this approach did not follow the Separation of Concern design pattern, which recommends breaking a computer program into distinct features that overlap in functionality as little as possible. In this case, in order to separate the parsing (step 1) from the building (step 2) processes, an intermediate model representation (Abstract Model Representation (AMR) in Figure 2) was required.

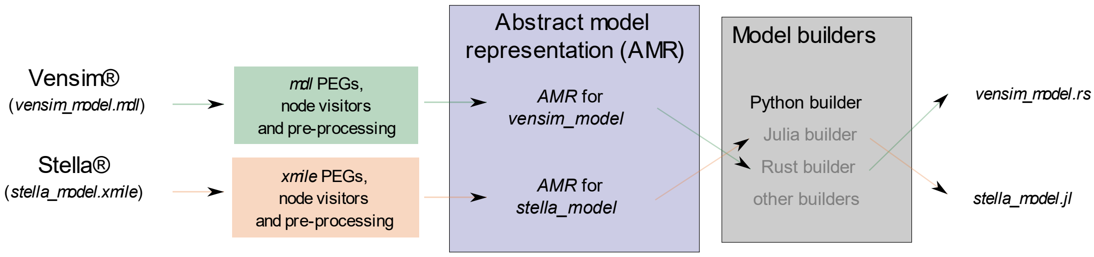


Figure 2: Translation workflow (Martin-Martinez et al., 2022)**.**

The AMR contains all the information included in the original model and adds the potential benefit of allowing the creation of builders for other programming languages.

This new translation workflow was introduced in version 3.0.0 and described in detail in (Martin-Martinez et al., 2022) and in the library documentation (https://pysd.readthedocs.io/en/master/structure/vensim_translation.html?highlight=supported%20Vensim#translation-workflow).

Release 3.0.0 marked a critical milestone for the future of the library, as it defines a clear structure to add new functionality and encourages contributions from the community.

2.1.3. ADDING NEW FEATURES AND OPTIONS

Among the most relevant new features included in PySD in the current Task are:

- Better support and handling of subscripts.
- Dynamic import of external data during model initialisation.
- A fully featured command line interface (CLI): this allows using all PySD functionality directly from the command line, without the need of writing all instructions on a Python script.
- The possibility to translate the code of different Vensim® views in separate Python modules and submodules: this results in a translated model which is more structured, organised and easier to read.
- Possibility to run modules independently: this was an additional benefit from the previous point, which could be used to validate different parts of the entire model individually.
- Possibility to store intermediate simulation results and restart a simulation from that point.
- Serialising external data and using it to initialize the model: the data that is read from spreadsheet files may be stored in binary format (netCDF), which allows for faster initialisation of the model.
- Adding a progress bar to see how much time is left until the simulation ends.
- Benchmarking tools for testing and comparing outputs between different files.
- Storing simulation results in three different file formats (.csv, .tab and .nc).
- A stepper function, to run simulations one step at a time (in progress)

2.1.4. IMPROVING RESULTING CODE CLARITY

The objective of the changes made in this dimension was to increase the readability of the Python code after translation and to facilitate its maintenance. As a result, translating a Vensim® model with the latest version of PySD results in fewer, cleaner and more organised lines of code.

Indeed, following a consistent naming convention for all Vensim® views in a model facilitates the organisation of the resulting Python code into modules and submodules. For instance, for a Vensim® model with three separate views (e.g., view_1, view_2 and view_3), setting *split_views* argument of the *read_vensim* function to True, creates the following tree inside the directory where the *.mdl* model is located (Figure 3):

```
main-folder
├── modules_many_views_model
│   ├── _modules.json
│   ├── view_1.py
│   ├── view_2.py
│   └── view_3.py
├── _subscripts_many_views_model.json
├── many_views_model.py
```

Figure 3: File structure resulting from setting the split_views argument to True in the read_vensim function for a model consisting of three separate views (i.e., view_1, view_2 and view_3). Adapted from https://pysd.readthedocs.io/en/master/advanced_usage.html#splitting-vensim-views-in-separate-python-files-modules.

In the previous figure, we see that the translation resulted in two files (*many_views_model.py* and *_subscripts_many_views_model.json*) and one folder (*modules_many_views_model*) inside the folder where he Vensim® model was located (*main-folder*). All the translated model equations are spread across each of the Python modules located in the *modules_many_views_model* folder, which are also named after the corresponding Vensim® views. This folder also contains the *_modules.json* file, which is a text file (in JSON format) which indicates the names of the variables included in each module. The *_subscripts_many_views_model.json* is also a JSON file, which includes all model subscripts. On the other hand, the *many_views_model.py* is the main model file that needs to be called to run the model and is named after the original Vensim® model. This file only includes the imports of all required libraries to run the whole model, the definition of control variables (initial_time, final_time, saveper, time_step), a function to load the contents of the two JSON files, and another function to load all modules inside the *modules_many_views_model* folder.

Figure 3 shows a very simple example of the structure of the translation of a simple model. However, in larger models, modellers may include a particular character in the View name to indicate that what comes after it is the name of the subview. For instance, one could name one Vensim® view as *ENERGY.Supply* and another one as *ENERGY.Demand*. In that case, setting the *subview_sep* keyword argument equal to [“.”] in the *read_vensim* method, would name the translated views as demand.py and supply.py and place them inside the ENERGY folder.

As can be seen from all the previous, this new feature results in a much cleaner and organised Python model.

Another improvement that has led to a significant reduction on the number of lines of code, is the possibility to load external data (from spreadsheet files) dynamically at model initialisation. In previous PySD versions this possibility did not exist, hence a preprocessing step was needed to hardcode all datasets of the spreadsheet files in the model file.

The docstrings of each function have also been simplified, and now provide the same information with fewer lines of code. Additionally, the description, the units and associated dimensions of each model element are available in the form of a pandas DataFrame by calling the *doc* method of the *Model* class.

Finally, the rewrite and restructuring the code described in Section 2.1.2, and especially the introduction of the AMR has also contributed to a reduction of the number of lines of code of the models translated with PySD.

2.1.5. IMPROVING USER FEEDBACK AND ERROR MESSAGES

Warnings and error messages have been added to the PySD library. The new error and warning messages have been very useful for identifying bugs in the Vensim® version of the WILIAM model. In most cases, these issues were not reported by Vensim®, and led to unexpected model results. Therefore, PySD acts as an additional layer of code quality assurance, regardless of whether the resulting Python code is going to be used or not.

The most relevant warning messages were used to warn the user in the following situations:

- When a variable is defined in more than one Vensim® view, when a control variable appears in a view or when a variable does not appear in any view as a view.
- When a subscript range is duplicated in a variable reference.
- When a translated variable has several types or subtypes.
- When trying to translate a function not supported by PySD.
- When extrapolating below the minimum value of above the maximum value of a lookup table at runtime.

The most relevant error messages would stop code execution in the following situations:

- When using Python versions older than 3.7
- When using a different version of PySD that was used to translate the model.
- When Vensim® parser fails.
- When issues occur while reading input parameters from spreadsheet files.

- When trying to run a model that includes a function not supported by PySD.
- When there is a dimension mismatch between x and y dimensions of lookups, or when the x includes a repeated value.
- When there is a circular initialization issue in the model.
- When the resource to allocate becomes negative in the ALLOCATE AVAILABLE and ALLOCATE BY PRIORITY functions.

All in all, the previous warning and error messages saved a lot of time during the debugging process.

2.1.6. IMPROVING TEST COVERAGE

More than 70 test models have been built to test the new functionality included in PySD. Some of which were included in the PySD library itself (https://github.com/SDXorg/pysd/tree/master/tests/more-tests), and some were included in a separate repository (https://github.com/SDXorg/test-models), which may be used by other System Dynamics projects.

In addition to the new test models, hundreds of unit tests have been added in the library which have left the test coverage of the library to an outstanding 100% (https://coveralls.io/github/SDXorg/pysd?branch=master). This implies that, in the worst-case scenario, there is at least one unit test validating the functionality of each function in the code.

2.1.7. FIXING BUGS

More than 70 issues reported in the issue tracker of the library (https://github.com/SDXorg/pysd/issues) by us and by other users of PySD have been fixed and closed. By including the bug fixes identified and fixed without reporting them in the issue tracker, the total number would significantly increase.

2.1.8. INCREASING EXECUTION PERFORMANCE

The redesign of the library has led to runtime performance improvements of up to 60%. For instance, the MEDEAS model used to run in around 45 minutes on a modern desktop computer with earlier versions of PySD, and it now runs under 15 minutes. The Python version of the WILIAM model runs in ~30 minutes, which is faster than the uncompiled Vensim® code, which takes more than 2 hours to run.

These performance improvements have mostly been achieved with the definition of a dependency dictionary, which creates more direct and efficient call stacks and with the reduction of boilerplate code. Other improvements have been achieved by optimising the implementation of specific functions. An example of this is the IF THEN ELSE function, which now only evaluates the branch for which the condition is satisfied.

The sdCloud certification report (https://sdcloud.io/reports/certificationReport/#), developed by the open-source System Dynamics community, may be used to check the performance improvements obtained with release 3.0.0 as compared to previous releases for several models (Figure 4).

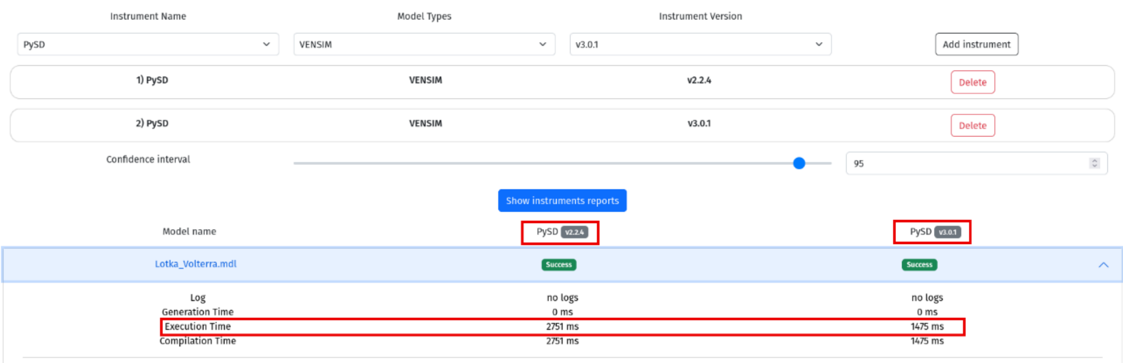


Figure 4: Execution time comparison for the Lotka Volterra model between versions 2.2.4 and 3.0.1 of PySD.

2.1.9. IMPROVING DOCUMENTATION

The new features and functions included in PySD during the present Task have been thoroughly documented. In addition, several existing sections of the documentation were revised, and some others were added to describe the new structure of the library and the new translation workflow. Finally, the what's new section was also included to detail the changes included in each new release of the library. The full documentation may be accessed in this link: https://pysd.readthedocs.io/en/master/.

2.1.10. OTHER RELEVANT CHANGES

Other relevant changes included in PySD that were made to give more visibility to the library, with the final goal of growing the community, were:

- Migration of the Github repository from the personal account of the original author to the SDXorg organization repository (https://github.com/SDXorg).
- Submitting the library to the conda-forge channel (https://conda-forge.org/).
- Improvements in the Continuous Integration (testing for different Python versions, Automated dependency updates, etc.).
- Updating the README file and creating a logo for the library.
- Publishing a paper about the library in the Journal of Open Source Software (Martin-Martinez et al., 2022).

2.2. TRANSLATION OF THE WILIAM MODEL

The improvements made on the PySD library, which were covered in the previous section, aimed at reaching a state that would allow to translate the model automatically, with minimum human intervention. Indeed, provided that both Python and the latest version of PySD are already installed in the system, only the following four lines of code are required in order to make the translation (Figure 5):

```python
import pysd
import sys

sys.setrecursionlimit(1500)

model = pysd.read_vensim("WILIAM.mdl",
                         initialize=False,
                         split_views=True,
                         subview_sep=["-", "."])
```

Figure 5: Lines of Python code required to translate the WILIAM model.

This creates a file structure like that of Figure 3, but specific to the modules and submodules of the WILIAM model.

However, as explained in Section 2.1.5: Improving user feedback and error messages, PySD also alerted of issues in the code that Vensim® does not report to the user, but that may result in unexpected simulation outcomes. To prevent these issues from accumulating, we run the translation of the WILIAM model with all new code introduced during each monthly development cycle and reported the identified issues to the modelling team. This way, the next development cycle would start with a bug-free version of the code.

The most common problems identified during the translation of these intermediate WILIAM versions were:

- Variables not belonging to any view or defined several times in a single view
- Resource to allocate being negative
- Values repeated in the x dimension of a LOOKUP or DATA table
- Dimension mismatch between the x and y dimensions of LOOKUP and DATA tables
- View names with wrong naming convention
- Repeated dimension in variable definitions
- Case sensitivity issues when calling spreadsheet files names, tabs or cell range names from within the model
- Invalid values inside cell range names in the spreadsheet files (e.g., string instead of number)
- Use of non-supported Vensim® (those reported before September 2022 were implemented in PySD)
- Circular dependencies

All in all, and until the time of writing, we reported over 60 issues to the teams in charge of the WILIAM model development, which has contributed to improve the quality of the final product, both in Vensim® and in Python.

The current internal release of the *pywiliam* model is based on WILIAM 1.1, which was released (also internally) at the end of July of 2023.

3. THE PYWILIAM INTERFACE

The *pywiliam* interface is an extra layer of Python code around the *pywiliam* model, to simplify the parametrisation of the model, to launch simulations and to visualise simulation results. The code is based on that developed for the MEDEAS model, but with numerous adaptations and improvements. The new code is currently not open to the public, but it will eventually be released under the MIT licence.

It includes a CLI with extra parametrisation options on top of those provided by PySD and a Graphical User Interface (GUI) to load and plot the simulation results.

To run a default simulation with the *pywiliam* model, the user can execute the *run.py* script without passing any additional arguments (Figure 6):

```
python run.py
```

Figure 6: Code required to run a default simulation with *pywiliam*.

The available options of the CLI (Figure 8) may be viewed by passing the *-h* argument to the main Python script, as follows (Figure 7):

```
python run.py -h
```

Figure 7: Command to view the *pywiliam* CLI options.

```
usage: usage: pywiliam [-h] [-v] [-n FILE] [-e FILE] [-p] [-c SHEET] [-x FILE] [-b] [-s]
                [-F VALUE] [-T VALUE] [-S VALUE]
                [--missing-values {warning,raise,ignore,keep}]
                [variable=new_value ...] [variable:initial_value ...]

WILIAM model

positional arguments:
  variable=new_value    redefine the value of variable with new value.variable must be a
                        model component, new_value can be a float or a a list of two list
  variable:initial_value
                        redefine the initial value of variable.variable must be a model
                        stateful element, initial_value must be a float

options:
  -h, --help            show this help message and exit
  -v, --version         show program's version number and exit
  -n FILE, --fname FILE
                        name of the results file, default is results_{scenario
                        sheet}_{initial time}_{final time}_{time step}.csv
  -e FILE, --externals FILE
                        path to the netCDF file where the external objects are stored
  -p, --plot            opens the plot gui after simulation
  -c SHEET, --scen SHEET
                        scenario file path
  -x FILE, --export FILE
                        export stateful objects states to a pickle at the end of the
                        simulation
  -b, --headless        headless mode (only CLI, no GUI)
  -s, --silent          silent mode. No user input will be required during execution.
                        Usefulwhen running batch simulations

model arguments:
  Modify model control variables.

  -F VALUE, --final-time VALUE
                        modify final year of the simulation, default is 2050.0
  -T VALUE, --time-step VALUE
                        modify time step (in years) of the simulation, default is 0.25
  -S VALUE, --saveper VALUE
                        modify time step (in years) of the output, default is 1.0 year

warning and errors arguments:
  Modify warning and errors management.

  --missing-values {warning,raise,ignore,keep}
                        exception with missing values, 'warning' (default) shows a warning
                        message and interpolates the values, 'raise' raises an error,
                        'ignore' interpolates the values without showing anything, 'keep'
                        keeps the missing values
```

Figure 8: Command line options of the *pywiliam* interface.

The *pywiliam* interface also enforces some sensible default values for several PySD options. For instance, the results are stored in a binary file (netCDF) rather than on a tab separated file (.tab), which is PySD's default output format. This way, the time spent writing the results to a file is reduced dramatically. In addition, the results file is placed in the *results* folder inside the project folder, and the file is named using the following convention:

results_SCENARIO-NAME_INITIAL-TIME_FINAL-TIME_TIME-STEP.nc

where SCENARIO-NAME is the name of the file where the scenario is defined, INITIAL-TIME, FINAL-TIME and TIME-STEP are the initial and final times of the simulation and the simulation time-step, respectively.

On the other hand, the Plot Tool (*plot_tool.py* module) is a GUI, which can be executed either in standalone mode (i.e., *python plot_tool.py*) or it can be loaded automatically when a simulation ends by passing the *-p* argument to the CLI (i.e., python run.py -p). The Plot Tool can load and plot simulation results exported from Vensim® or obtained from a Python translation of a Vensim® model. It supports the following file formats: csv, tab and nc. It can plot multidimensional time-series and may be used to compare time-series from multiple different scenarios.

Figure 9 shows the design of the Plot Tool. The File menu on the top left corner is used to load simulation results files. Right below, there is the panel where all the variables found in the results files are listed in alphabetical order. On top of the list, there is a search bar, which facilitates the location of specific variable names among the whole list. When a variable is selected, the corresponding time-series values are displayed on the right panel. Plots may also be exported to *png* files using the save button, on top of the plot, and the description of the variable may be viewed by clicking on the information button.

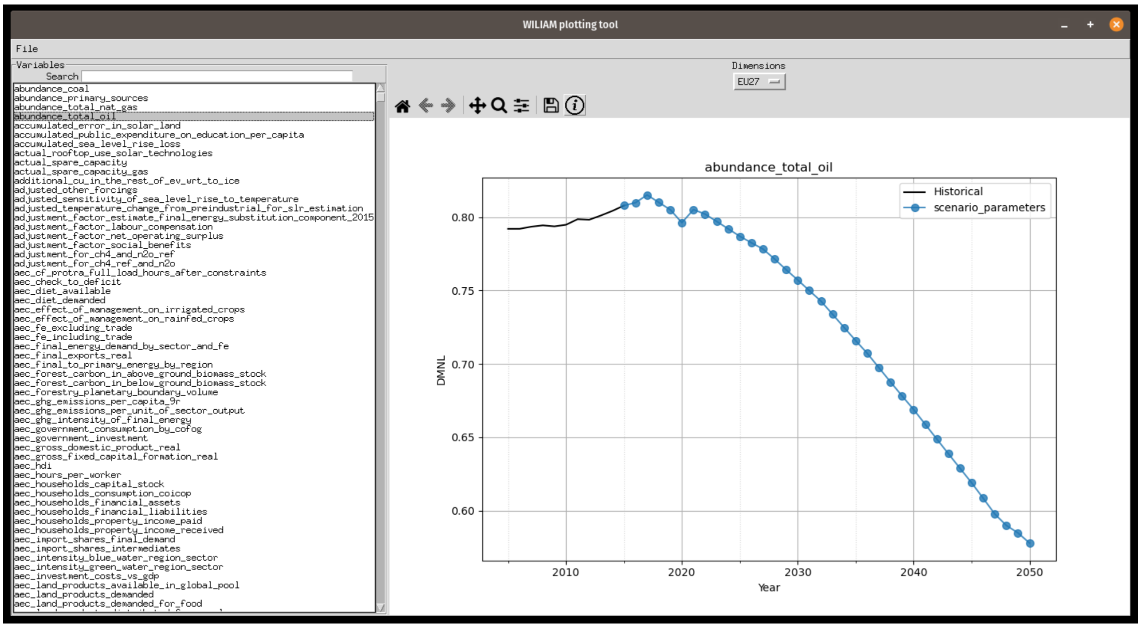
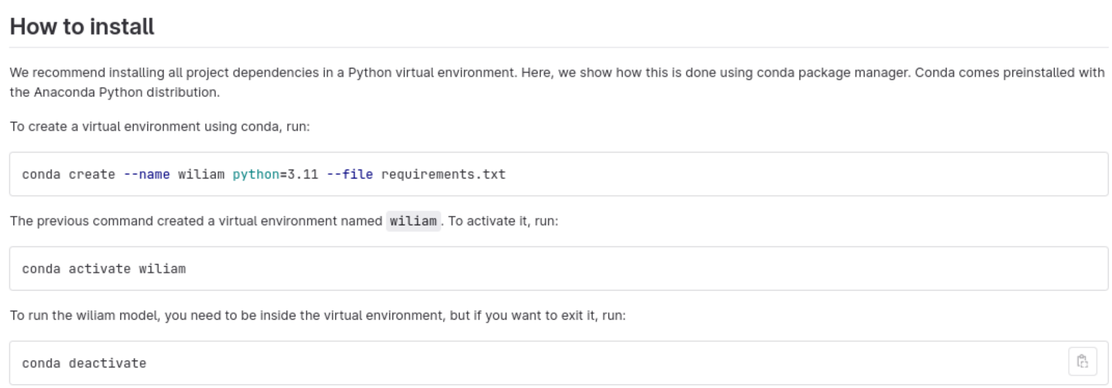


Figure 9: The *pywiliam* Plot Tool GUI.

On the other hand, although the contents of the simulation results (netCDF file) can be visualised with the Plot Tool, an additional Jupyter Notebook is included in the code, to allow for exporting datasets to csv or tab files. The package also includes a test suite, which guarantees that the code behaves as expected.

Finally, the README.md file included in the repository also provides installation and operation instructions (Figure 10).



How to install

We recommend installing all project dependencies in a Python virtual environment. Here, we show how this is done using conda package manager. Conda comes preinstalled with the Anaconda Python distribution.

To create a virtual environment using conda, run:

```
conda create --name wiliam python=3.11 --file requirements.txt
```

The previous command created a virtual environment named `wiliam`. To activate it, run:

```
conda activate wiliam
```

To run the wiliam model, you need to be inside the virtual environment, but if you want to exit it, run:

```
conda deactivate
```

Figure 10: Installation instructions in the README file of the *pywiliam* interface repository.

4. THE WILIAM E-HANDBOOK

The WILIAM e-handbook was envisioned as a short document introducing the WILIAM model, its main purpose, and the conceptual models and hypothesis included in each of its modules. The e-handbook is not meant to become the full documentation of the WILIAM model, but instead is expected to provide a quick glance of its main features. This document will eventually be shared in workshops and dissemination activities and may be exported in different formats (pdf, epub, html, etc).

The present Task has led the writing of the e-handbook, which has been a coordinated effort of most partners of the LOCOMOTION consortium. In order to produce a homogeneous and harmonised document, in this Task we created a table of contents of the e-handbook, as well as a template with the basic structure and the main topics that all chapters corresponding to the different WILIAM modules should have in common. Once the different partners shared their contributions, they were reviewed, combined and homogenised, and some introductory and linking sections and figures were included to give it the final structure.

The contents of the e-handbook are included hereafter:

4.1. INTEGRATED ASSESMENT MODELS

The growing awareness of climate change and its link to the essential energy transition has become an important issue in our society. Therefore, scientists, policy makers and intergovernmental institutions are now dedicating their efforts on how to deal with these significant challenges. To assist in decision-making to address these issues, scientists from a wide range of fields, from physics and chemistry to economics, engineering or sociology, collaborate to develop Integrated Assessment Models (IAMs).

An IAM is a numerical simulation tool designed to help understand the relationships between many technological, economic, environmental and social variables that characterize the development of our society. Models built from historical data allow you to simulate future scenarios with different alternatives of action to guide decision-making. In recent years, these models have been applied to the search for alternative solutions to climate change and the energy transition. To address these complex problems, an IAM requires integrated knowledge from a wide range of areas: climatology, economics, engineering, sociology or politics, seeking to represent the interactions between human beings and the environment.

IAMs began to develop in the 1970s with the pioneering World3, developed by a team led by Donella and Dennis Meadows and from which the report "The Limits of Growth" was obtained (Meadows et al., 1972). Over the next few years, and as the power of computers increased, new IAMs such as DICE (Nordhaus, 1979), IMAGE (Rotmans et al., 1990), GCAM (Edmonds & Reilly, 1985), MESSAGE (Messner & Strubegger, 1995), MERGE (Manne & Richels, 2005), and WITCH (Bosetti et al., 2006) emerged. As new models were being developed, new features were also added. There are currently dozens of IAMs in the literature with very different approaches that assist in decision-making at all institutional levels.

However, although their approaches and evolution have been very different, most IAMs have similar structures. They are usually structured in different modules that correspond to the different dimensions represented in the model (economy, energy, climate, land use, etc.). There are relationships and feedbacks between the different modules of the model that make the characteristics of one module influence the others. As can be seen in Figure 11 (Evans & Housefather, 2020), different inputs affect these modules. These inputs usually correspond to different hypotheses and assumptions made, such as population or GDP. Policies are also inputs in the model. To the right of the figure are the outputs, which are the results obtained in the different modules of the model once the policies have been applied.

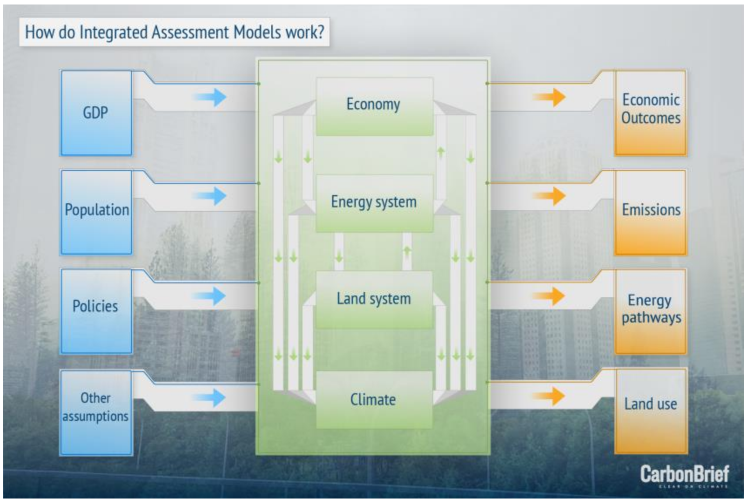


Figure 11: Basic structure of operation of IAMs (Evans & Housefather, 2020).

It is important to note in relation to the outputs and results provided by IAMs, that the objective of such models is not to make an accurate forecast of the different variables in the future, but to provide viable representations of what may happen in the future based on a coherent set of assumptions and hypotheses.

The information provided by an IAM presents many uncertainties, both parametric and structural, which has generated criticism and controversy throughout history. The interpretation of the results of IAMs should be accompanied by uncertainty analysis tools that narrow the margins of such uncertainty in probabilistic terms. Another basic requirement for interpreting the results of IAMs is the transparency and substantiation of the assumptions made. To this end, in recent decades, a series of international scenarios or narratives have been developed that attempt to unify hypotheses and allow viable descriptions of how the future could develop under different socioeconomic, technological and environmental conditions. Most models currently use the scenarios generated through the "Shared Socio-Economic Pathways" (SSPs) (O'Neill et al., 2014). These narratives, based on demographic, economic, technological, and even lifestyle change assumptions, enable a better analysis of mitigation and adaptation policies. However, uncertainty is not the only limitation that IAMs have. Some of the criticisms that IAMs receive are that there is a great lack of transparency in many of the models, the relationships and feedbacks between modules are often very limited or simplified, especially in the case of climate change (Diaz & Moore, 2017; Evans & Housefather, 2020), or the models tend to focus on the economic-technological part, largely forgetting the social part and human behavior (Barnes et al., 2013; van den Berg et al., 2019). Section 4.1.2 will discuss the limitations of IAMs.

Despite these limitations, the models have already proven useful in assessing the magnitude of the problem of climate change and the effectiveness of possible solutions (Weyant, 2017). A good proof of this is that they are the tools used by the main institutions to develop and evaluate their policies. For example, the IPCC uses multi-model studies, such as AMPERE (Kriegler et al., 2015) or EMF27 (Blanford et al., 2014) and the European Commission uses the POLES model (Després et al., 2018).

4.1.1. IAMS CLASSIFICATION

The difficulty in representing the relationships between energy, economy and environment together with uncertainty in the assumptions made in the models, leads to the existence of very different approaches in the development and structure of IAMs.

One of the most widespread classifications in the literature is what Weyant does in his contribution to the state of IAMs (Weyant, 2017). It defines the profit-cost models as those simple models that provide a very aggregate representation of climate change mitigation costs. They generally use simplified equations and have been used to calculate the optimal trajectory of global greenhouse gas (GHG) emissions, and the corresponding prices to be charged for these emissions. These models do not show the detailed relationships between energy, economy and environment. Examples of these models are DICE (Nordhaus, 1979), FUND (Anthoff & Tol, 2014), and PAGE (Hope, 2011). On the other hand, Weyant defines detailed process models as those complex models that seek to provide more detailed projections of the effects of climate change using economic issues and projections of physical effects, such as the reduction of energy resources, lands flooded by rising sea levels, etc. This vision of IAMs as more complex models is the one that best fits a multidisciplinary view of the relationships between human behavior and its effects on the environment. Examples of these models are GCAM (Edmonds & Reilly, 1985), IMAGE (Rotmans et al., 1990), and MESSAGE (Messner & Strubegger, 1995).

Other widespread classifications of models in the literature are the proposals in the revision of IAMs of (Capellán-Pérez, 2016). In his study he classified the main existing models according to the following criteria: the method of implementing policies, the level of complexity of systems, the level of integration between subsystems (feedbacks), the treatment of uncertainty, the economic balance and the treatment of energy in the economy (Capellán-Pérez, 2016). It is important to note that each of these modelling criteria has its own strengths and weaknesses, making it difficult to conceive an integrated model capable of providing the best solutions to all issues.

The classifications of the IAMs shown so far are quite "theoretical", but when studying the models and analyzing the results obtained in them, as is done in IPCC multi-studies, it is necessary to classify them according to other more "technical" aspects. In these studies, models are classified according to:

- The type of license: all models should have a license, regardless of whether they are intended for public use (open source) or not.
- The programming language: models use different languages, such as Python, GAMS, Vensim®, or Visual Basic, in their development.
- The flexibility of the model: This shows the extent to which predefined conditions can be changed in the model, for example: reallocating capital across sectors, how easily the economy can replace energy technologies, or whether there are limitations on fossil fuels and renewable resources.
- The forecasting method: In perfect forecasting models, all future decisions are accounted for with current decisions. In contrast, recursive-dynamic (myopic) models make decisions at any given time based only on information from that time period.
- The level of geographic aggregation: models can be developed for one or more regions or be global models.
- The temporal dimension: models differ in their base year, their time step and their time horizon.

Finally, models are also differentiated in the inputs that drive the model (drivers), the way they represent the different sectors and systems, and by the type of policies applied. Each model, being developed with a different approach and with very different objectives, elaborates some parts of the system in more detail than others. In order to compare the different models in all these issues, the consortium of IAM developers (IAMC) has created a wiki where the main features of many models are explained (https://www.iamcdocumentation.eu/index.php/IAMC_wiki).

4.1.2. IAMS LIMITATIONS

It is important to note that, in the field of IAMs, existing criticisms and limitations are constructively understood and help to improve models' development. The limitations of IAMs are still significant and of very different types. They can be gathered into four large groups:

First, there is a distinct lack of transparency in many of the models; some of these models could be considered "black boxes" with non-existent or difficult-to-understand documentation. In this sense, many models are developed with a private license, which makes it very difficult to review their internal equations. The lack of transparency in how models are developed, and functioning makes them very difficult for policymakers, who are not used to getting involved in modelling. This, coupled with the lack of communication with developers, makes IAMs seem distant and complicated for policymakers.

Second, although IAMs are composed of different modules, they are usually characterized by a fairly sequential structure and the existing feedbacks between the different modules that structure the model are scarce. Particularly relevant is the case with the feedback from the effects of climate change. Models that do consider this feedback often do so through damage functions. However, as (Diaz & Moore, 2017) show, damage functions have many limitations. An example of this is that they consider temperature increases of up to 4 or 5ºC at the end of the century, with a negligible representation of impacts on the economy. In this sense, realizing the effects of climate change on the regions and sectors that suffer the most from their impacts is a formidable challenge for IAMs (Weyant, 2017). The feedback from land-use modules with energy modules to explore climate mitigation scenarios under diverse political and technological conditions is also very limited. Although the shortage of feedback on the environmental part is the most relevant, there is also a lack of interconnectivity between the other modules. A large percentage of IAMs lack, for example, relationships that allow the economic impact of energy policies to be analyzed, the effect of mineral resource scarcity on the economy, or the consequences of changes in society's behavior. Especially striking is the case that population and GDPpc are exogenous in most models.

Third, IAMs often focus on the economic-technological part, forgetting in many cases human behavior and largely the social part (Barnes et al., 2013). For IAMs, capturing human behavior is arguably the most difficult of their goals, and models representing consumer preferences are very scarce (van den Berg et al., 2019). Most of the decisions made in the models are from an economic point of view, even though this is not the only point of view that society uses to make its decisions. Generally, the welfare implications associated with these decisions have not been thoroughly evaluated in the literature.

Finally, most models follow a very similar line of thought, consider similar hypotheses, the same technologies and achieve very similar results. Plurality in hypotheses and scenarios is very sparse, leading to proposals which are not generally very innovative. One of the most common hypotheses in the literature is that the economy follows the model of a perfect market through optimization methods and the perfect substitution of factors, as well as widespread use of prices as indicators of scarcity (Scrieciu et al., 2013). Another of the most repeated issues in IAMs is the use of negative emissions policies. Many models consider such options as bioenergy carbon capture and storage (BECCS) to reduce total emissions. However, these options are presently not applicable, and uncertainty about their use is very high, as neither economic nor energy viability has been proven (Bednar et al., 2019). The abundance of fossil fuels is a default assumption in most IAMs. However, this assumption is questioned by studies in the literature showing that fossil fuel extraction could face significant constraints in the coming decades (Capellán-Pérez et al., 2016). In general, there are always the same assumptions, so it is necessary to expand the range of options to explore, especially those outside the "mainstream".

By applying similar models and scenarios, a large part of the models achieves fairly similar results in terms of emission reduction, GDP or energy resource developments. Alternative models and scenarios developed outside major research groups have a more difficult scientific and social impact. All these limitations stimulate the continuous development of existing models. IAMs aim to keep pace with the development of sector-specific models and to fill the gaps that remain in the literature.

4.2. THE WILIAM MODEL

WILIAM is a system dynamics policy-simulation model, descendent from the MEDEAS model (Capellán-Pérez et al., 2020) which has been designed to explore long-term decarbonization pathways within planetary boundaries by addressing a series of limitations of existing IAMs. In fact, despite the high number of IAMs, many share several relevant and disputable hypotheses/characteristics. The main shortcomings aimed to be addressed with WILIAM are:

- Lack of plurality/simplified representation of economic processes typically based on optimization.
- Equilibrium dynamics.
- Aggregate production functions and representative agents (Hardt & O'Neill, 2017; Scrieciu et al., 2013).
- Future energy transitions modelled as demand-driven transformations (assumption of future high energy availability at affordable cost, both for renewables and non-renewables).
- The neglect of implications of future energy investments required to achieve the transition to renewables for the entire system Energy Return on Energy Investment of the full system (Capellán-Pérez et al., 2019).
- Difficulties to reach 100% renewable systems.
- Underestimation of the damages caused by climate change.
- The absence of the material dimension and key sustainability dimensions other than climate change.

WILIAM focuses on the detailed representation of the economic processes following a Dynamic Econometric Input-Output approach and consistently linking the economic and biophysical spheres according to the principles of Ecological Macroeconomics. WILIAM follows a complex system approach, in which the interactions between dimensions are more relevant than the complexity within each module. System Dynamics allows to capture complex feedback loops and nonlinear relationships among social, economic, and environmental variables.

WILIAM comprises 8 modules of Earth and human systems: (1) Demography, (2) Society, (3) Economy, (4) Finance, (5) Energy, (6) Materials, (7) Land and Water, and (8) Climate. Figure 12 shows the structure overview with the main linkages between modules. Different modules reach different levels of detail and complexity. WILIAM starts to run in 2005 and typically runs until 2060, although the simulation horizon may be extended to 2100. WILIAM is a multiregional model which blends top-down and bottom-up (end-use) modelling approaches, and it integrates knowledge and methods from different disciplines aiming to capture the main dynamics between human and natural systems. This comprehensive integration is particularly relevant when modelling disruptive scenarios involving significant shifts in economic structure, social values, norms, and individual behaviour. Indeed, the ultimate goal of WILIAM is to explore the social, economic, and environmental implications at the global and regional levels of long-term socio-ecological transition pathways, considering biophysical planetary limits as well as socio-economic constraints. The validation of the model has been carried out following several of the usual validation procedures of models in system dynamics (Barlas, 1996; Sterman, 2000). The historical data has been used for a first validation, and the results will also be compared with other models.

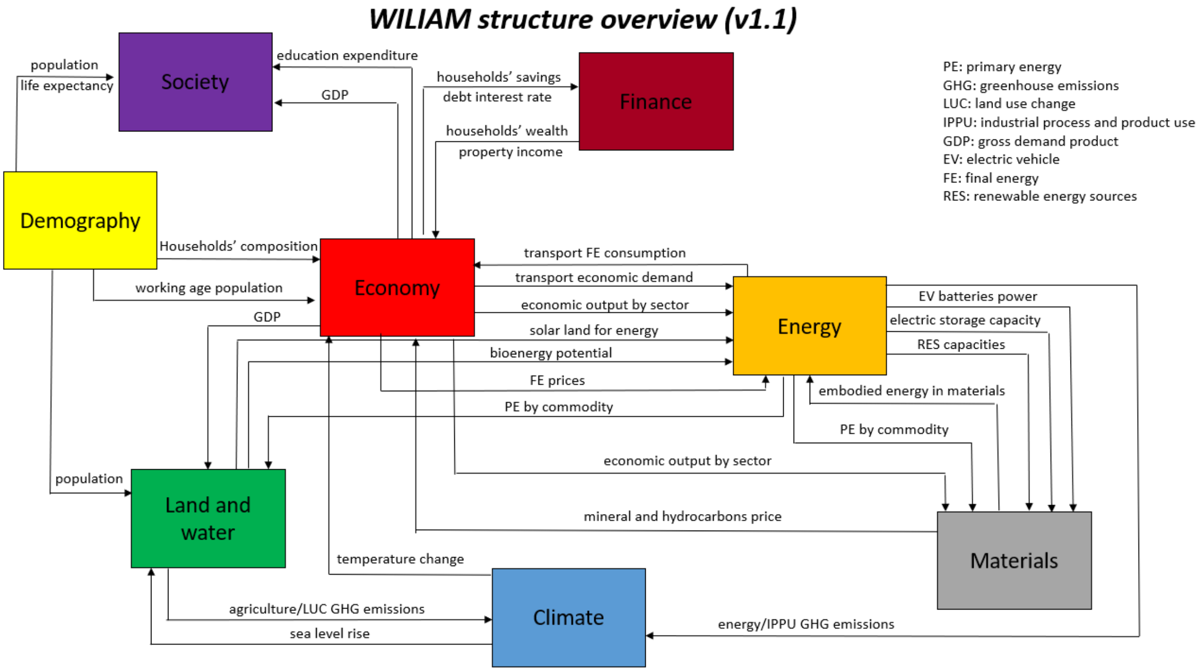


Figure 12: WILIAM (v1.1) simplified structure overview. Main linkages between modules.

Below is a summary of the main characteristics of WILIAM model:

- Open source (MIT license).
- Modular structure.
- Detailed geographical coverage: multi-regional world model with 8 global regions and the integration of the 27 EU countries individually for some dimensions.
- Modules for demography and society, which allow to consider feedbacks such as migration and the effects of climate change on population.
- The economy is represented by a dynamic econometric model covering 35 regions, with detailed representation of consumption (including 60 household's types), production (based on input-output tables of 62 sectors), government (including collection of taxes + public expenditures), investment, labour, international trade and financial dimensions.
- Transport: a detailed representation of passenger transport including 11 transport modes, 10 power trains and a portfolio of behavioural policies.
- Energy supply: representation with high detail of the full energy supply-chain through the main refinery, transformation, and supply processes.
- Variability of renewables keeps track of sub-annual time scale effects on annual energy balances depending on the current power system setup of build-up of generation capacities and flexibility capacities (demand-side management, storage, sector coupling, hydrogen and synthetic fuels).
- Computation of the EROI of the full system considering the EROI and material requirements of green technologies, which feedbacks the energy demand.
- Techno-sustainable potentials of renewables considering biophysical, geographical, natural resources and EROI constraints.
- Modules of fossil fuels and metals fully integrated with Energy and Economy modules. The models are driven by a demand-price mechanism in which a higher price 1) reduces the demand for materials, and 2) triggers investment to increase the material extraction capacity, limited by available resources and increasing depletion rates.

- Land-use module including human food, energy, and climate interactions, thus allowing endogenizing land-based renewable potentials (solar and bioenergy), considering the agriculture and land related emissions, and the effects of climate change on biophysical variables such as crop yields.
- Water module: main outputs are water availability and water stress based on demand and supply.
- Climate module converting emissions coming from the other modules in changes in main climate variables, such as mean temperature change and sea level rise.
- Climate change impacts on capital stock and labour productivity.
- The improvement in scenario assessment by integrating demand management policies across modules, particularly for passenger transport and diets.
- A rich portfolio of conventional and heterodox policies (CO_2 taxes, basic income, behavioural changes, working time reduction, recycling, etc.) aiming at being able to simulate a broad range of narratives, such as Green Growth, Green Deal and Post-growth/Degrowth.

4.3. The Demography module

4.3.1. GENERAL DESCRIPTION

The economy is conditioned by humans, including their consumption. Food, energy, extraction of resources and their use are derived from the composition, size location and characteristics of the population. The population also produces the workforce for the economy, therefore an improved understanding and endogenization of age of population enables modelling of labour. This section briefly explains the physical dimension, i.e., mathematical equations representing the sex (female, male) and age of humans (5-year cohorts until 80) by region (35), including annual migration flows, fertility and mortality.

A conceptual scheme of the Demography module is shown in Figure 13. Endogenous variables in the population dynamics are births, deaths, and migration flows, expanded to the life expectancy at birth and mortality rates. Exogenous data feed the model with historical data, which is also used to implement realistic scenario hypothesis for the future trends. These are:

- According to the literature, women are the drivers of reproduction, being those affected by both fertility rates and gender ratio to estimate new people in the model.
- An exercise of endogenization has been the calculation of mortality rates (by sex, age cohort and region) as of life expectancy at birth (by sex and region) through exponential equations, i.e., one by age cohort. The analytic equations are reproduced in the model with the same parameters, constant assumptions.
- Migration flows are complex and there is not a comprehensive method in the literature to represent the causes and effects on demographic changes. Consequently, the model applies exogenous assumptions as constant values that can be modified in case the user would like to test specific scenarios. Emigration rates get the number of people that leave from each region, then shares of emigration allocate such quantities across the rest of the regions with net balance equal zero.
- Some relationships have been considered from the literature about climate change impacts in relation to the loss/improvement in health, represented by the life expectancy at birth. The parameters of those relationships are exogenously assumed as constant hypothesis.
- Switches are a common feature the user may activate or not specific parts of the system. Specifically, there are two switches in the Demography module, one to facilitate the omission of migration flows and the other to activate the feedbacks to the life expectancy at birth.

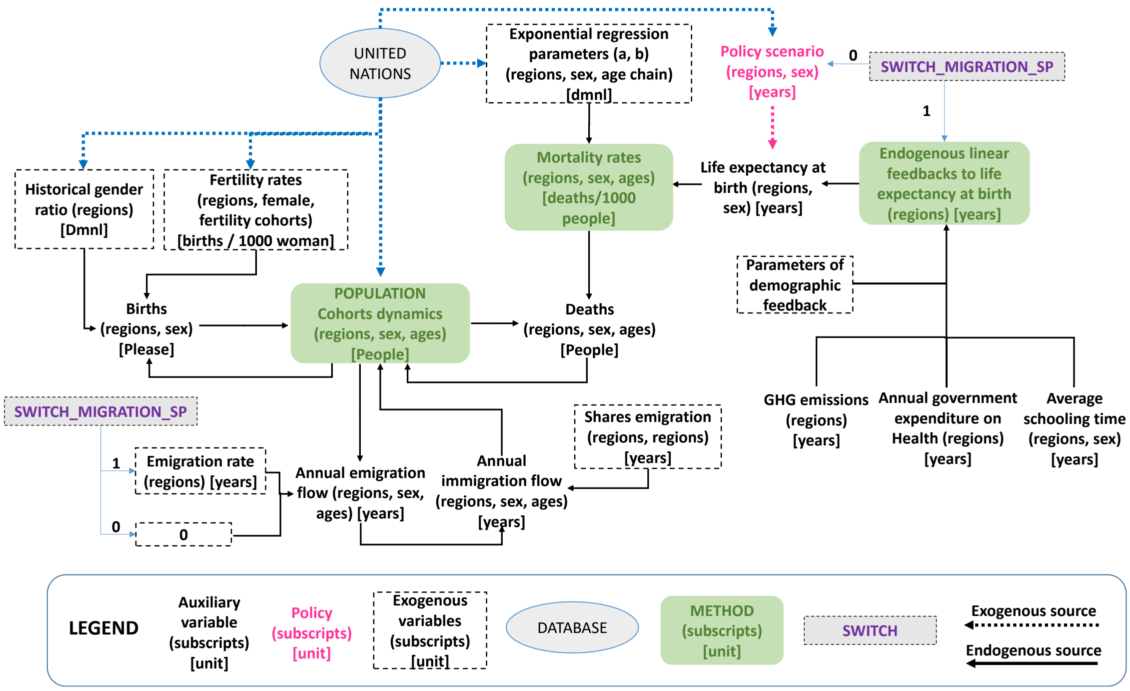


Figure 13: General scheme of the Demography module in WILIAM.

Interactions with other modules go in both directions:

- Climate (GHG emissions), economy (annual government expenditure on health), and the average schooling time (education) have influence in the life expectancy at birth.
- Population clearly affects the evolution of household's composition which, in turn, it has influence on consumption throughout demand functions in the economy.
- Also, in the economy, people available for the labour market come from the demography.

The demand of food is determined by diets, an approach presented in the chapter dedicated to the Land and Water module (Section 4.7). It is highlighted here that the nutritional assessment does not include feedback to demographic variables since there is not agreement in literature to do it at the level of aggregation of WILIAM.

4.3.2. METHODOLOGICAL APPROACH

The main stock of demography is mathematically represented with Equation 1 to Equation 14. The general dynamics consists of three parts, i.e., the initial stock (0-4 years old), the intermediate stock (0-80 years old), and the last stock with the remaining population (greater than 80 years old). The equations are different for each of them. Auxiliary variables are required to update all the stock each five years, timing variables based on modules and delays.

In the following equations, P_t is the population in time t. Δ means the "variation" of flow of people in a time step of the integration process. Em and Im represent the emigration and immigration flows, respectively. $P5y_t$ is the population that should be removed in the stock each 5 years. $AUXP_t$ is simply a delay (time step) of $P5y_t$. $EmRates$ are the emigration rates. $ImShares$ are the shares of immigration to allocate emigration from the region of origin to the rest. FR means fertility rate. GBR is the gender birth ratio. MR means mortality rate. LEAB is the acronym for life expectancy at birth. C_m are the coefficients of impacts to the life expectancy at birth. Finally, $\Delta effect$ is the annual variation of those dimensions considered to affect the life expectancy at birth.

The internal management of different stocks by age cohort needs the combination of continuous and discrete modelling:

$$P_t(Reg_i, Sex_j, Age_k) = \Delta P_t(Reg_i, Sex_j, Age_k) - Em\,(Reg_i, Sex_j, Age_k) + Im\,(Reg_i, Sex_j, Age_k)$$

Equation 1

$$\Delta P_t(Reg_i, Sex_j, Age_{0-4}) = \max(\,births_t(Reg_i, Sex_j, Age_{0-4}) - deaths_t(Reg_i, Sex_j, Age_{0-4}) + \frac{P5y_t(Reg_i, Sex_j, Age_k)}{5}, -\frac{P_t(Reg_i, Sex_j, Age_{0-4})}{t_s})$$

Equation 2

$$\Delta P_t(Reg_i, Sex_j, Age_{k:\,5-79}) = \max(-deaths_t(Reg_i, Sex_j, Age_k) + \frac{P5y_t(Reg_i, Sex_j, Age_{k-1})}{5}, -\frac{P_t(Reg_i, Sex_j, Age_k)}{t_s})$$

Equation 3

$$\Delta P_t(Reg_i, Sex_j, Age_{+80}) = -deaths_t(Reg_i, Sex_j, Age_{+80}) + \frac{P5y_t(Reg_i, Sex_j, Age_{75-79})}{5}$$

Equation 4

$$P5y_t(Reg_i, Sex_j, Age_k) = P_t(Reg_i, Sex_j, Age_k)\ \ when\ mod(t,5) = 0$$

Equation 5

$$P5y_t(Reg_i, Sex_j, Age_k) = AUXP_t(Reg_i, Sex_j, Age_k)\ \ when\ mod(t,5) \neq 0$$

Equation 6

$$AUXP_t(Reg_i, Sex_j, Age_k) = P5y_{t-t_s}(Reg_i, Sex_j, Age_k)$$

Equation 7

The other two flows are related to the phenomena of migration:

$$Em\,(Reg_i, Sex_j, Age_k) = P_t(Reg_i, Sex_j, Age_k)\ \cdot EmRates\,(Reg_i)$$

Equation 8

$$Im\,(Reg_i, Sex_j, Age_k) = \sum_{i=1}^{35} Em\,(Reg_i, Sex_j, Age_k)\ \cdot ImShares(Reg_i, Sex_j, Age_k)$$

Equation 9

Births are calculated with two exogenous indicators, birth ratio to decide the sex, and fertility rates to know from which female cohorts' births came.

$$births\,(Reg_i, Sex_{female}, Age_{0-4}) = \frac{\sum_{k=15}^{49} P_t(Reg_i, Sex_{female}, Age_{k:15-49})\ \cdot FR(Reg_i, Sex_{female}, Age_{k:15-49})}{2} \cdot (2 - GBR(Reg_i\,))$$

Equation 10

$$births\ (Reg_i, Sex_{male}, Age_{0-4}) = \frac{\sum_{k=15}^{49} P_t(Reg_i, Sex_{female}, Age_{k:15-49})\ \cdot FR(Reg_i, Sex_{female}, Age_{k:15-49})}{2} \cdot GBR(Reg_i\ )$$

Equation 11

Finally, deaths have an expanded method to life expectancy at birth, variable being affected by other parts of the model.

$$deaths\ (Reg_i, Sex_j, Age_k) = \ P_t(Reg_i, Sex_j, Age_k)\ \cdot MR(Reg_i, Sex_j, Age_k)$$

Equation 12

$$MR(Reg_i, Sex_j, Age_k) = \ A(Sex_j, Age_k)\ \cdot\ e^{B(Sex_j, Age_k) \cdot LEAB(Reg_i, Sex_j) \cdot}$$

Equation 13

$$\frac{LEAB(Reg_i, Sex_j)}{dt} = \ C_m \cdot \Delta effect_m$$

Equation 14

4.3.3. DATA SOURCES

All the data used in this module has been gathered from the United Nations database (United Nations, 2023).

4.3.4. IMPLEMENTED POLICIES

The Demography module is useful to observe future trends of population distribution and size. A user can customize and apply three predefined policies:

- Fertility rates: this policy target defines the fertility rates for the future in regions by 2050. Values are based on the historical period (2005-2020). SELECTION: low fertility rates (1, left box in sheet "demography_data"), average (2, box in the middle in sheet "demography_data"), or high (3, right box in sheet "demography_data").
- Life expectancy at birth: this policy target defines the life expectancy at birth for the future in regions. Values are based on the historical period (2005-2020). SELECTION: minimum LEAB (1), average (2), or maximum (3). In the model there is a SWITCH (SWITCH DEM LIFE EXPECTANCY AT BIRTH) to choose between (0) exogenous pathway and (1) endogenous feedbacks for life expectancy at birth.
- Migrations: this policy target activates (1) or disactivate (0) the existence of international bilateral migration flows for the future. Default included data from (Abel & Cohen, 2022), but they may be changed.

4.3.5. KEY OUTPUTS

The main outputs of the Demography module are:

- Births and deaths.
- Net emigration flows by region.
- Stock of population and derived aggregations

4.4. THE ECONOMY MODULE

4.4.1. GENERAL DESCRIPTION

The core of the Economy module of WILIAM is an input-output (IO) model that has been extended by endogenous final demand (as in a type ii IO model) and by mutual feedbacks between quantities and prices (as in a Computable General Equilibrium (CGE) model) (Figure 14). In contrast to CGE models, the IO model in WILIAM has New Keynesian features, as markets are not generally cleared by the price mechanism, but effective demand under supply constraints determines the outcome for the different industries. The macroeconomic IO model in WILIAM is especially designed for incorporating feedbacks between the economy and nature. That comprises primary energy supply, land use and water supply as well as the feedbacks of climate change to the economy.

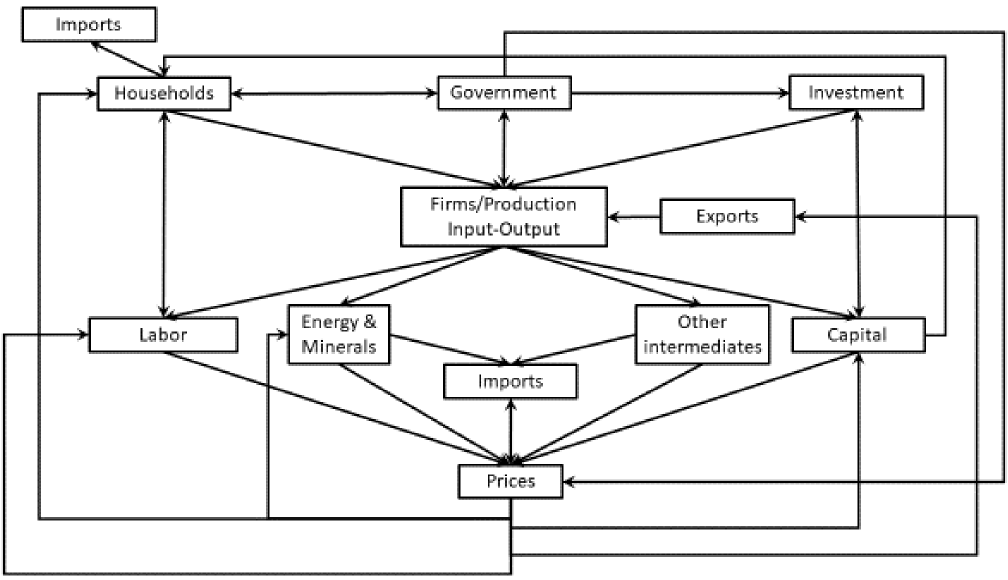


Figure 14: Main components and relations of the economy module.

4.4.2. SUBMODULES

The Economy module of WILIAM encompasses eight submodules: households, government, firms/production, energy and materials, investment/capital, labour, international trade and prices.

The Households submodule of WILIAM covers multiple household types that consume and react asymmetrically depending on their economic circumstances and preferences over time. The submodule is inspired on the architecture of the consumption block of the FIDELIO model (Kratena et al., 2017) and create a structure of demand equations in which 13 different bundles of goods are distinguished. Households consume according to their preferences, out of disposable income and out of their real assets and financial situation.

On the other hand, it is well known that IAMs have been widely used to assess technology-based solutions to curb global warming, but moving people's lifestyles towards climate-friendly behaviour could have an enormous potential as well (van de Ven et al., 2018). Given the relevance of behavioural change in climate change mitigation, some alternatives for its modelling have been developed in WILIAM.

The Government submodule comprises the main interactions among households, firms and the General Government, which encompasses Central Government, State Government, Local Government and Social Security Funds. The core of the module is based on the System of National Accounts (United Nations, 2010)[1] and covers the different components of the government revenue, expenditure and balance (deficit/surplus). The different components of the government accounts are linked to other parts of the economic model through a number of equations describing the relations between the different agents of the economy. The architecture of this submodule is inspired in the government module of the FIDELIO model (Kratena et al., 2017), but it has been modified in several directions. First, it has been adapted to operate in a System Dynamics framework by introducing lagged variables and modifying the mechanism of adjustment of revenue-expenditure-deficit, in order to eliminate the simultaneity of some equations. Second, the model has been enriched by including a better description of the government final consumption expenditure and gross fixed capital formation following the Classification of the Functions of Government (COFOG). Third, tax rates on income and wealth have been modelled in order to capture the progressivity of fiscal systems. Finally, the model has been linked to and is consistent with the Government Financial Accounts.

The Production submodule is Dynamic Econometric Multi-regional IO model. The IO model employed in WILIAM belongs to the class of Keynesian macroeconomic IO models, where behaviour of the economic agents (e.g., industries, households) is based on econometric estimations which are derived mainly from New-Keynesian economic theories with some elements of traditional neoclassical microeconomics (substitution in trade, consumer demand systems). The multi-regional input-output (MRIO) framework is also used for the price IO model, where each region chooses shares of imports in total expenditure (for intermediate and final users) for given composite good prices. The latter is a function of all regional output prices. Among the new features of the production block are the specification of the gross fixed capital formation and the dynamics of capital stock; the new price submodel; the new submodule of international trade; the new labour submodule; the integration with the bottom-up Land and Water, Energy and Materials modules; or the integration of climate change impacts and adaptation. The number of regions and industries represented in the model has also been expanded, and special attention has been paid to the representation of those industries more relevant for the analysis of decarbonisation pathways such as primary industries, energy sector or transportation. The latter has been specifically conceived to ensure a better connexion with several modules of the WILIAM, especially the land, the materials and the energy modules. The link between the Economy, Energy and Materials modules of WILIAM has been reinforced in WILIAM increasing the consistency of the model and introducing dynamism in the IO structure.

The International trade submodule describes the substitution potential in trade flows, triggered by factors that determine substitution between imported and domestic products on the one hand and between different sources of origin of imported products on the other hand. The MRIO framework in WILIAM comprises trade flows information at the detailed level of users, i. e. of industries and final users (private and public consumption, investment) for each receiving region (by sending region and user). Trade elasticities are introduced into this framework via flexible expenditure shares of each user for each good from any region. Substitution in trade via Armington or gravity equations is introduced at two different nests for every industry and final private consumption: (i) domestic/imported, and (ii) imports by origin. The former interacts with regional IO coefficients of the MRIO system. The general technology in terms of total IO coefficients is given for the International trade submodule. This is determined in other parts of the model, mainly by structural change within industries (electricity sector) and by the productivity of aggregate inputs (labour, energy and intermediates). For any given set of total IO coefficients of a region, the regional coefficients of the MRIO system change with trade substitution. This is the main characteristic of trade substitution in the MRIO framework, namely the feedback on regional output patterns.

The Labour market submodule of WILIAM is not specified as fully competitive like in CGE models with full employment, where the price of labour balances supply and demand. Unemployment in these models can only be voluntary and is caused by reactions of the labour supply to the wage rate or by the distortion

[1] The website of EUROSTAT offers an overview of the National Accounts.

of the price of labour due to payroll taxation. In these models, the price of labour exerts an influence both on the supply and the demand of labour. The latter is usually described by a labour demand function derived from a production or cost function (in most cases CES). The supply of labour reacts with a certain elasticity to the wage rate without - in most cases - differentiating between the intensive (hours) and the extensive (persons) margin.

In WILIAM, the price mechanism is not trimmed to balance all markets towards a general equilibrium. Though, it incorporates a price feedback from the labour market (wage bargaining) affecting the supply curve via a price mechanism. The price equation also includes an adjustment mechanism that ensures that demand does not exceed maximum production capacity. Beside these feedbacks, sectoral outputs are in principle demand driven. In contrast to CGE models, WILIAM is motivated by a short to medium-run perspective and thus explicitly capture the transition between the past and the future (non-equilibrium states), whereas CGE models compare two different full equilibrium states, i.e., before and after a system intervention takes place (long-run perspective).

The regional output price equation is formulated as a mark-up on variable costs, i.e., unit costs of intermediates and labour unit costs. The mark-up, therefore, covers the capital cost term as well as a mark-up in the narrow sense due to imperfect markets. It is specified as a function of the industry output gap and therefore will take into account the positive impact of capacity utilisation on the rate of return on capital as well as inflation expectations that depend on the output gap for a given monetary policy.

4.4.3. METHODOLOGICAL APPROACH

This section describes the equations of the different submodules of the economic module. The main elements of the equations include endogenous variables, exogenous parameters (denoted by the prefix EXO_) and econometric parameters (denoted by the prefix Constant_, Alpha_, Beta_, Gamma, …). All variables are expressed in nominal terms expect those with the suffix _real, which indicates that the variable is expressed in 2015 constant prices. The suffixes _BP and _PP indicates that the variables are expressed in basic and purchaser prices respectively. Additionally, the following sub-scripts are used:

- q, r, s: denote regions.
- t: denotes year.
- u: represents all the industries (i) and final demand components (c).
- i, j: represent all the 62 industries in the WILIAM IO framework. We can distinguish four types of industries :
  - k: non-energy industries
  - Energy industries[2]:
    - f: industries selling final energy to industries and households (i.e., mining of coal and lignite, extraction of peat; petroleum refinery; transmission, distribution and trade of electricity; manufacture of gas, distribution of gaseous fuels through mains and steam and hot water supply).
    - q: industries transforming energy inputs into final energy (i.e., petroleum refinery; production of electricity by coal; production of electricity by gas; production of electricity by nuclear; production of electricity by hydro; production of electricity by wind; production of electricity by petroleum and other oil derivatives; production of electricity by solar photovoltaic; production of electricity by solar thermal; production of electricity n.e.c., including production of electricity by biomass and waste, production of electricity by tide, wave, ocean

[2] Note that some of energy-related industries belong to more than a group. This is the case for petroleum refinery which belongs at the same time to group e, f and t. It is also important to take into consideration that here we do not address the fact that some sectors extract and/or produce energy materials which can be used for both energy and non-energy purposes (i.e., petroleum refinery to produce diesel, gasoline, kerosene, etc. but also plastics).

and production of electricity by geothermal; manufacture of gas, distribution of gaseous fuels through mains and steam and hot water supply).

  - e: energy-extracting industries and industries supplying energy for transformation (i.e., mining of coal and lignite; extraction of peat; extraction of crude petroleum and services related to crude oil extraction, excluding surveying; extraction of natural gas and services related to natural gas extraction, excluding surveying; extraction, liquefaction, and regasification of other petroleum and gaseous materials; mining of uranium and thorium ores; petroleum refinery; manufacture of gas, distribution of gaseous fuels through mains and steam and hot water supply).

- c: represents the components of the final (c): household consumption, consumptions of Non-profit institutions serving household, government consumption, gross fixed capital formation, change in inventories and valuables.
- l: represents the 13 household consumption categories.

4.4.3.1. THE HOUSEHOLDS SUBMODULE

In MEDEAS, households' demand depended exclusively on their main source of income, wages (Capellán-Pérez et al., 2020). Nevertheless, other income sources or even wealth portfolios may affect households' total final consumption expenditure. This issue can only be analysed in an approach that incorporates household heterogeneity, either by explicitly dealing with different household types or – if the data do not allow it – by a specification that implicitly considers household heterogeneity.

In the macroeconomic literature of the last decades, the most important aspect of household heterogeneity refers to the consumption reaction to transitory income changes. These differences in consumption propensities with respect to income and net assets explain differences in fiscal multipliers. In disaggregated models with input-output cores – either CGE or Macro-Econometric (ME) models – the heterogeneity of households implies that different income-to-consumption reactions plus socio-demographic characteristics like age can have an impact on the economic structure and the aggregate economic outcomes (Kim et al., 2013). Macroeconomic models have introduced different household types in terms of their reaction in consumption to transitory income shocks. That led to the distinction between "Current Income" (CI) and "Permanent Income Hypothesis" (PIH) households as – for example – in (Campbell & Mankiw, 1991) and (Auerbach & Gorodnichenko, 2012) or to the distinction between "savers" and "borrowers" as in (Eggertsson & Krugman, 2012). One line of research that explains empirical deviations from the PIH is the "buffer stock model" (C. D. Carroll, 1996; José Luengo-Prado, 2006). This strain of the literature emphasises income uncertainty and down payments for durables as specific savings motives that explain deviations in consumption behaviour from the pure PIH model. In addition, previous literature notes that total resources of households – "cash on hand" – are the major driver of households' consumption in both micro- and macro-economic applications (Jappelli & Pistaferri, 2014; Kratena et al., 2017). The concept of cash on hand generally measures total household resources as the sum of disposable income, net financial assets and housing stock. For modelling households in WILIAM we follow a similar philosophy.

There are two blocks of equations in the household submodule. On the one hand, resource equations determine households' disposable resources according to their economic situation. On the other hand, demand equations determine how much households spend on different goods and services depending on their resources and preferences. Households are modelled in all regions of WILIAM i.e., the EU-27 countries plus eight non-EU regions. Nevertheless, our approach distinguishes two different casuistries according to the number of household types (H) considered in each region. While the two approaches differ in the number of households considered, they are modelled using the same equations (both in structure and components), which ensures comparability of results across regions. In the following, we explain the main differences between the two approaches:

Household heterogeneity in a multiple households approach ($H > 1$).

In these regions, households are explicitly modelled from a heterogeneous perspective, i.e., the number of household types considered is larger than one. In general, a total of 60 household typologies are modelled, representing the population by groups of interest, i.e., by urban/rural location, household composition and income level. Such a heterogeneity enables the model to capture macroeconomic (Fisher et al., 2020; Krueger et al., 2016), distributional (Böhringer et al., 2019) and environmental (Sommer & Kratena, 2017) asymmetry of effects, stemming from the existence of different propensities to consume across household typologies and countries (C. Carroll & Kimball, 1995; Jappelli & Pistaferri, 2014, 2020)). Moreover, in these regions, we account for changes in the weights of different types of households over time, which makes it possible to analyse, for example, the effects of changes in demographic composition (e.g., urbanisation), socio-economic shocks (e.g., increasing income inequality) or shifts towards climate-friendly behaviour (e.g., more households with environmentally sustainable consumption habits).

Household heterogeneity in a single household approach ($H = 1$).

Unlike the previous approach, for these regions the modelling of households is less sophisticated and is based on a homogeneous approach, i.e., households are represented by a single representative household (the average household in society). For these cases, therefore, heterogeneous effects among households are not distinguished directly, but can be taken into account by assuming that only one share of total households reacts like CI households (Campbell & Mankiw, 1991) by taking that into account in the aggregate consumption function, as laid down in (Sarantis & Stewart, 2003). These demand equations reflect the average reactions of consumers to shifts in their economic situation and preferences.

In what follows, we describe the equations through which households make their consumption decisions, depending not only on disposable income but also on their economic stability reflected in terms of wealth (buffered by indebtedness) or the evolution of products' prices.

4.4.3.1.1. HOUSEHOLD INCOME AND WEALTH

Households receive primary incomes from different sources that, once redistributed, determine their disposable resources either for consumption or saving. Thus, households' disposable income ($Disposable\_income_{rht}$) of a generic household h (with $h = 1, \dots, H$) from region r (with $r = 1, \dots, R$) in a given year t (with $t = 1, \dots, T$) is defined through the following accounting identity:

$$HH\_Disposable\_income_{rht} = \frac{1}{Number\_households_{rht}} \times \{\textstyle\sum_i [Share\_Labour\_income_{rhit} \times (1 - Exo\_rate\_Social\_security_{rit}) \times Labour\_compensation_{rit} + Share\_Operating\_surplus_{rhit} \times (1 - Exo\_Tax\_rate\_Operating\_Surplus_{rit}) \times Gross\_Operating\_surplus_{rit}] + Exo\_HH\_Assets\_yield_{rht} \times HH\_Assets_{rht} - Exo\_Tax\_rate\_Income_{rht} \times HH\_Income_{rht-1} - Exo\_Tax\_rate\_Wealth_{rht} \times Net\_wealth_{rht-1} + Share\_Social\_benefits_{rht} \times Social\_benefits_{rt} + Exo\_Other\_transfers_{rht}\}$$

Equation 15

where,

- $Number\_HH_{rht}$ is the number of housholds by type and region.
- $Share\_Labour\_income_{rhit}$ $Share\_Capital\_income_{rhit}$, $Share\_Social\_benefits_{rht}$ are the share of Labour income ($Labour\_compensation_{rit}$), operating surplus ($Gross\_Operating\_surplus_{rit}$), and social benefits ($Social\_benefits_{rt}$) received by each household.
- $Exo\_rate\_Social\_security_{rit}$ is the rate of social contributions.

- $Exo\_Tax\_rate\_Operating\_Surplus_{rit}$, $Exo\_Tax\_rate\_Income_{rht}$, and $Exo\_Tax\_rate\_Wealth_{rht}$. are the taxz rates on the operating surplus, household income and wealth respectively.
- $Exo\_HH\_Assets\_yield_{rht}$ is the yield of the assets owned by each household (HH_Assets_{rht})
- HH_Income_{rht} is the gross income *by household* and is calculated as:

$$HH\_Income_{rht-1} = \sum_i [Share\_Labour\_income\_by\_HH_{rhit-1} \times (1 - Exo\_rate\_Social\_security_{rit-1}) \times Labour\_compensation_{rit-1} + Share\_Capital\_income\_by\_HH_{rhit-1} \times (1 - Exo\_Tax\_rate\_Operating\_Surplus_{rit-1}) \times Gross\_Operating\_Surplus_{rit-1}] + Exo\_HH\_Assets\_yield_{rht-1} \times HH\_Assets_{rht-1} + Share\_Social\_benefits\_by\_HH_{rht} \times Social\_benefits_{rt} + Share\_Transfers\_by\_HH_{rht} \times Transfers_{rht} + Exo\_Other\_transfers_{rht}$$

Equation 16

where,

- Net_wealth_{rht} is the net wealth by households computed in the financial module
- $Exo\_Other\_transfers_{rht}$ are other transfers received by households.

4.4.3.1.2. HOUSEHOLD CONSUMPTION

On the consumption side, one of WILIAM's priorities is to model the heterogeneity of the propensity to consume of different types of households in different macroeconomic environments, depending on their preferences and available resources (disposable income and net wealth). Accordingly, the equations presented below aim at capturing – as much as possible – how such a heterogeneity across households affects the demand for goods. We deal with different dimensions of heterogeneity, as consumption is undertaken by several types of households, on the one hand, and split up into distinct categories and goods, on the other hand. This is accomplished by means of multiple interlinked equations that take into account the heterogeneous effects across households in different regions and, at the same time, also allow us to determine both total consumption and its distribution across goods (Figure 15).

At the top level of Figure 15, we have a distinction between durable and non-durable goods. According to (Sarantis & Stewart, 2003), we can expect a relatively stable proportion of durable to non-durable expenditure in the mid-term. Therefore, total household consumption expenditure ($Consumption\_total_{hrt}$) can be obtained as the sum of durable ($Consumption\_durable_{hrt}$) and non-durable ($Consumption\_non\_durable_{hrt}$) consumption expenditure:

$$Consumption\_Total_{rht} = Consumption\_Durable_{rht} + Consumption\_Non\_Durable_{rht}$$

Equation 17

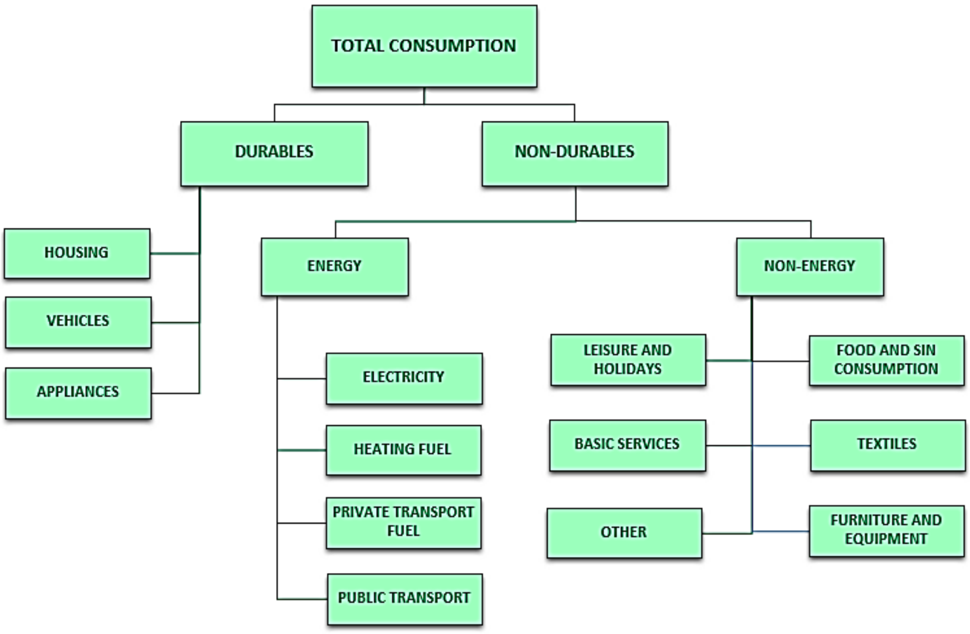


Figure 15: Overview of household consumption in WILIAM.

4.4.3.1.2.1. CONSUMPTION OF DURABLES

There are three types of durable goods: housing services ($Consumption\_Housing\_services_{rht}$), appliances ($Consumption\_Appliances_{rht}$) and vehicles ($Consumption\_Vehicles_{rht}$), so that the consumption of durables is determined as follows:

$$Consumption\_Durable_{rht} = (Propensity\_Consumption\_Housing\_services_{rht} + Propensity\_Consumption\_Appliances_{rht} + Propensity\_Consumption\_Vehicles_{rht}) \times Disposable\_income_{hrt}$$

Equation 18

The consumption of each of the components of durable consumption is calculated by multiplying their respective propensities to consume times the disposable income:

$$Consumption\_Housing\_services_{rht} = Propensity\_Consumption\_Housing\_services_{rht} \times Disposable\_income_{hrt}$$

Equation 19

$$Consumption\_Appliances_{rht} = Propensity\_Consumption\_Appliances_{rht} \times Disposable\_income_{hrt}$$

Equation 20

$$Consumption\_Vehicles_{rht} = Propensity\_Consumption\_Vehicles_{rht} \times Disposable\_income_{hrt}$$

Equation 21

Housing services include actual and imputed rents, as well as other costs for the maintenance and use of the dwelling. In this regard, it is worth clarifying why imputed rents are included as part of households' consumption expenditure to avoid misinterpretation. National Accounts treat the rental value of the tenant-occupied housing and the imputed rental value of owner-occupied housing as an expenditure. This implies accounting for a hypothetical rent that owners would have paid if they were renting their housing to themselves. The underlying reason for that is to ensure comparability of macroeconomic measures over time and across countries with no distortions associated with fluctuating rental market activity. Moreover, housing services must be distinguished from investment in housing, considered gross fixed capital formation instead of consumption. Keeping these insights in mind, households' propensity to consume of housing services is obtained as follows:

$$\log(Propensity\_Consumption\_Housing\_services_{rht}) = Constant\_Housing\_services_{rh} + Beta\_Housing\_services\_1_{rh} \times log(Disposable\_income_{hrt}) + Beta\_Housing\_services\_2_{rh} \times log[Delta\_buffer_{rht} \times (Housing\_stock_{rht} + Financial\_assets\_stock_{rht})]$$

Equation 22

where $Constant\_Housing\_services_{rh}$, $Beta\_Housing\_services\_1_{rh}$, $Beta\_Housing\_services\_2_{rh}$ and $Delta\_buffer_{rh}$ are region- and household type-specific parameters. This specification is similar to that of appliances and vehicles. The first three are derived from econometric estimates and have the following interpretations: $Constant\_Housing\_services_{rh}$ is a constant that captures different region and household type effects; $Beta\_Housing\_services\_1_{rh}$ is the income elasticity and $Beta\_Housing\_services\_2_{rh}$ stand for the wealth effect on expenditure (i.e., the effect whereby households' consumption grows as the value of their assets – both dwelling and financial portfolios – increase). Wealth accumulation generates confidence in households about their present and future economic stability encouraging consumption, but this effect is softened by indebtedness. Hence, the parameter $Delta\_buffer_{rht} = (Housing\_stock_{rht} + Financial\_assets\_stock_{rht})/(Housing\_stock_{rht} + Financial\_assets\_stock_{rht} - Loans\_stock_{rht})$ serves as a buffer to the wealth effect as households' debts increase in relation to their assets.

4.4.3.1.2.2. CONSUMPTION OF NON-DURABLES

The consumption of non-durables (i.e., all those items excluded from the durable equations presented above) is determined as follows:

$$log(Consumption\_Non\_Durable_{rht}) = Constant\_Non\_Durable_{rh} + Beta\_NonDurable\_1_{rh} \times log\,(Disposable\_income_{rht}) + Beta\_NonDurable\_2_{rh} \times log[Delta\_buffer_{rht} \times (Housing\_stock_{rht} + Financial\_assets\_stock_{rht})]$$

Equation 23

As shown in Figure 15, non-durables are split up between energy and non-energy commodities. On the one hand, energy goods are modelled in a very detailed way by incorporating specific equations for different energy demands and connecting them with the energy module.

For all transport related expenditures, i.e., the sum of fuel expenditure for private transport ($Consumption\_Private\_transport_{rht}$) and expenditure in public transport ($Consumption\_Public\_transport_{rht}$), consumption expenditure is derived in three steps. First, total demand for transportation in million passenger-kilometres (pkm) ($Total\_PassengerKm_{hrt}$) is obtained from the following equation:

$$log(Total\_PassengerKm_{rht}) = Constant\_Total\_PassengerKm_{rh} + Beta\_Total\_PassengerKm\_1_{rh} \times log\left(\frac{Consumption\_Non\_Durable_{rht}}{Price\_Non\_Durable_{hrt}}\right) + Beta\_Total\_PassengerKm\_2_{rh} \times log(Price\_Transport_{rht})$$

Equation 24

where $Constant\_Total\_PassengerKm_{rh}$, $Beta\_Total\_PassengerKm\_1_{rh}$, $Beta\_Total\_PassengerKm\_2_{rh}$ are region- and household type-specific parameters obtained from econometric estimates that have the following interpretations: $Constant\_Total\_PassengerKm_{rh}$ is a constant that captures different region and household type effects; $Beta\_Total\_PassengerKm\_1_{rh}$, measures the elasticity of total pkm with respect to total expenditure on non-durables, being $Price\_Non\_Durable_{rht}$ a price index for these products; and $Beta\_Total\_PassengerKm\_2_{rh}$ reports the price elasticity of demand for pkm. Note that $Price\_Transport_{rht}$ is a price derived from the weighted sum of the prices of each m mode of transport, such that:

$$Price\_Transport_{rht} = \sum_m Share\_Transport\_mode_{rhmt} \times Price\_Transport\_mode_{rhmt}$$

Equation 25

where $Share\_Transport\_mode_{rhmt}$ refers to the shares of each mode m of transport in total pkm demand and $Price\_Transport\_mode_{rhmt}$ denotes prices per pkm for each of the modes of transport[3].

The total transportation demand is passed to the bottom-up transportation module which computes, among other variables, the demand of public transportation by mode ($PassengerKm\_Public\_transport_{rhmt}$, in million pkm) and the physical energy use for private transportation by fuel ($Phy\_Energy\_Private\_transport_{rhft}$, in TJ). These variables are then used by the economic module to compute the expenditure in fuel for private transportation and the expenditure in public transportation. The fuel input by type for private transportation coming from the transport submodule ($Phy\_Energy\_Private\_transport_{rhftf}$), is transformed into expenditure in current prices ($Consumption\_Private\_Transport_{rhft}$, in million \$) using a set of implicit prices that convert physical energy use into real expenditure ($Price\_Implicit\_Private\_transport_{rft}$, in $mill\ \$/TJ$) and the price indexes of the different fuels calculated in the production submodule ($Price\_Private\_transport_{rft}$):

$$Consumption\_Private\_transport_{rhft} = Phy\_Energy\_Private\_transport_{rhft} \times Price\_Implicit\_Private\_transport_{rft} \times Price\_Private\_transport_{rft}$$

Equation 26

Similarly, public transportation by mode ($PassengerKm\_Public\_Transport_{rhmt}$) coming from the transport submodule is converted in expenditure in current prices:

$$Consumption\_Public\_transport_{rhmt} = PassengerKm\_Public\_transport_{rhmt} \times Price\_Public\_transport_{rmt} \times Price\_Public\_transport_{rmt}$$

Equation 27

On the other hand, household energy demand for heating/cooling, lighting, appliances etc., is modelled at the level of the single energy categories f: solid fuels, liquid fuels, gas and electricity. The equations for these categories are analogous to the transport equation, but describe expenditure for energy as follows:

$$log(Consumption\_HH\_Energy_{rhft}) = Constant\_HH\_Energy_{rhft} + Beta\_HH\_Energy\_1_{rhf} \times log\left(\frac{Consumption\_Non\_Durable_{rht}}{Price\_NonDurable_{rht}}\right) + Beta\_HH\_Energy\_2_{rf} \times log(Price\_HH\_Energy_{rft}) + Beta\_HH\_Energy\_3_{rf} \times log(Efficiency\_HH\_Energy_{rft})$$

Equation 28

The price of the energy category ($Price\_HH\_Energy_{rft}$) and the efficiency of the building stock for heating ($Efficiency\_HH\_Energy_{rt}$) are both region-specific[4]. Energy expenditure in monetary units is transformed into physical units using implicit prices and prices from the production module.

From the link between the energy module and the household submodule, we can arrive at the amount of money spent on energy and transportation ($Consumption\_HH\_Energy_{rhft}$,

[3] For private transportation the price index refers to fuel price.

[4] Another option to deal with prices is combining the energy price and the efficiency variable into one “service” price term.

$Consumption\_Private\_transport_{rhft}$ and $Consumption\_Public\_transport_{hrmt}$). Thus, it is possible to derive the total energy expenditure ($Consumption\_Non\_Durable\_Energy_{rht}$) as the sum of all these parts:

$$Consumption\_Non\_Durable\_Energy_{hrt} = \sum_f Consumption\_HH\_Energy_{rhft} + \sum_f Consumption\_Private\_transport_{rhft} + \sum_m Consumption\_Public\_transport_{rhmt}$$

Equation 29

Given the assumption of full separability between energy and non-energy goods, the total household consumption of non-durable non-energy goods ($Consumption\_Non\_Durable\_Non\_Energy_{rht}$) can be drawn as a remnant as follows:

$$Consumption\_Non\_Durable\_Non\_Energy_{ht} = Consumption\_Non\_Durable_{rht} - Consumption\_Non\_Durable\_Energy_{rht}$$

Equation 30

These expenditures are distributed among six different non-durable non-energy bundles of goods. In order to allocate the total consumption of non-durable non-energy goods among the different bundles of goods considered, we construct a system of demand equations. Using this approach, we determine the share of a generic bundle of goods m in $Consumption\_Non\_Durable\_Non\_Energy_{rht}$. Hence, the total expenditure for a generic category of non-energy non-durable consumption ($Consumption\_Non\_Durable\_Non\_Energy\_by\_category_{rhmt}$) is as follows:

$$Consumption\_Non\_Durable\_Non\_Energy\_by\_category_{rhmt} = Consumption\_Non\_Durable\_Non\_Energy_{rht} \times Share\_QUAIDS_{rhmt}$$

Equation 31

$Consumption\_Non\_Durable\_Non\_Energy_{rht}$ (budget share) for each type of household and region. These weights explain the share of expenditure on each of the different goods as a function of the prices of all bundles of goods and total non-energy non-durable expenditure.

$$Share\_QUAIDS_{rhmt} = Constant\_QUAIDS_{rhm} + \sum_n [Beta\_QUAIDS\_1_{mn} \times log(Price\_QUAIDS\_category_{rhnt})] + Beta\_QUAIDS\_2_m \times \log\left(\frac{Consumption\_Non\_Durable\_Non\_Energy_{rht}}{Price\_QUAIDS\_1_{rht}}\right) + \frac{Beta\_Quaids\_3_m}{Price\_QUAIDS\_2_{rht}} \times \left[log\left(\frac{Consumption\_Non\_Durable\_Non\_Energy_{rht}}{Price\_QUAIDS\_1_{rht}}\right)\right]^2$$

Equation 32

where $Constant\_QUAIDS_{rhmt}$, $Beta\_QUAIDS\_1_{mn}$, $Beta\_QUAIDS\_2_m$ and $Beta\_QUAIDS\_3_m$ are terms derived from the estimation of a QUAIDS model (Banks et al., 1997) that have the following interpretations: $Constant\_QUAIDS_{rmnt}$ is a constant that captures region and household type effects on the shares of the different bundles of goods; $Beta\_QUAIDS\_1_{mn}$ reflects the effects on the budget share of changes in the own and cross-prices of the different bundles of goods m and n; and, $Beta\_QUAIDS\_2_m$ and n report the effects on the budget share of the change in households' expenditure on non-energy non-durable goods. Both $Price\_QUAIDS\_1_{rht}$ and $Price\_QUAIDS\_2_{rht}$ are price indices calculated as:

$$log(Price\_QUAIDS\_1_{rht}) = \sum_i Share\_QUAIDS_{rhit} \times log(Price\_QUAIDS\_category_i)$$

Equation 33

$$log(Price\_QUAIDS\_2_{rht}) = \prod_i (Price\_QUAIDS\_category_i)^{Beta\_QUAIDS\_2}$$

Equation 34

As mentioned above, the equations of the household submodule determine consumption in terms of representative households, either one ($H = 1$) or more than one ($H > 1$) depending on each region. Therefore, the vector of consumption expenditure by type of good obtained from the consumption equations for each household type and region must be multiplied by the number of households it represents out of the total households in society (i.e., by its population representativeness). These weights vary over time and are calculated based on information from the economic and socio-demographic modules. Thus, updates of the representativeness of the household types in each region are based on the following data. On the one hand, if $H > 1$, the number of households represented by each of our 60 household typologies varies as a function of demographic and/or economic changes affecting the variables used in their definition. In particular, the socio-demographic module provides, for each region, information on the number of households for 12 different typologies[5]. This information is combined with data on the distribution of income per household type coming from the economic module. Thus, depending on the economic situation and income redistribution, households will move between low- and high-income levels, affecting the demand for goods and services and welfare measures. On the other hand, if $H = 1$, we have only one representative household in the region and, therefore, the data needed is just the total number of households in these regions over the years simulated in WILIAM. This information comes from the demographic module.

The result of the system of equations described in this section is a set of vectors of consumption of 13 categories, for each type of representative households in each region in purchaser pieces $Consumption\_by\_category_{rhlt}$. In order to link these vectors to the IO model, a series of transformation are required. First, the consumption has to be multiplied by the number of households by category in order to get the aggregated consumption by region:

$$Consumption\_by\_category\_Aggregated_{rhlt} = Consumption\_by\_category_{rhlt} \times Number\_households_{rht}$$

Equation 35

The resulting vectors of aggregated consumption by 13 categories, household type and regions are transformed into vector of household final consumption expenditure by industry ($Final\_demand\_PP_{ri,HFCE,t}$, which is part of $Final\_demand\_PP_{rict}$) using a set of bridge matrices linking the 13 consumption categories and the 62 industries of WILIAM :

$$Final\_demand\_PP_{ri,HFCE,t} = Consumption\_by\_category_{rhlt} \times Number\_households_{rht}$$

Equation 36

This vector will be further transformed into a vector of final demand in real terms and basic prices $Final\_demand\_BP\_real_{ri,HFCE,t}$ (see section 4.4.3.3.6) and linked to the IO model .

4.4.3.2. THE GOVERNMENT SUBMODULE

The government submodule includes a set of equations which calculate the main components of the government accounts, including the government revenue, expenditure, and the balances and change in the stock of debt.

The functioning of the government submodule can be summarised in two stages:

[5] Urban single adult with children; urban single adult without children; urban couple with children; urban couple without children; other urban households with children; other urban households without children; rural single adult with children; rural single adult without children; rural couple with children; rural couple without children; other rural households with children; and, other rural households without children.

1. For every time step, the government makes a prediction of the evolution of nominal GDP and imposes some specific policy objectives for the total revenue to GDP and deficit/surplus to GDP ratios and (by difference) for total expenditure (all in nominal terms).
2. The actual revenue, expenditure, deficit/surplus and change in gross debt are calculated considering the exogenous parameters and the policy variable, as well as the evolution of different endogenous variables in the model.

Thus, the submodule starts with the assumption that the Government of each of the countries/regions of the model (r) forecasts, at the beginning of each time step (year), three key variables:

- the GDP growth for the incoming year: $GDP\_growth\_obj_{rt}$ (by the fault is equal to the GDP growth rate in the previous period)
- the objective value for the Deficit/Surplus to GDP ratio: $Gov\_DeficitSurplus\_to\_GDP\_obj_{rt}$,
- the objective ratio for the total Revenue to GDP ratio: $Gov\_Revenue\_to\_GDP\_obj_{rt}$,

These three variables are given exogenously (by default or changed by the user).

These parameters are used to compute government's objectives in terms of GDP, revenue and budget balance:

$$GDP\_obj_{rt} = GDP_{rt-1} \times [1 + Exo\_GDP\_growth\_obj_{rt}]$$

Equation 37

$$Gov\_Revenue\_obj_{rt} = Exo\_Gov\_Revenue\_to\_GDP\_obj_{rt} \times GDP\_obj_{rt}$$

Equation 38

$$Gov\_DeficitSurplus\_obj_{rt} = Exo\_Gov\_DeficitSurplus\_to\_GDP\_obj_{rt} \times GDP\_obj_{rt}$$

Equation 39

Once the target values for government deficit/surplus and revenue have been obtained, the total projected expenditure can be calculated as

$$Gov\_Expenditure\_obj_{rt} = Gov\_Revenue\_obj_{rt} - Gov\_DeficitSurplus\_obj_{rt}$$

Equation 40

This variable will be used to calculate some of the components of the government expenditure in Section 4.4.3.2.2.

4.4.3.2.1. GOVERNMENT REVENUE

Next, we describe the equations for the calculation of the different components of the government revenues.

The first component of the government revenues is the 'Taxes on products, receivable'. This component is linked to the intermediate and final demand of goods and services calculated in the consumption and production submodules. However, the link is not straightforward. First, the values reported in the IO model are in basic prices and, accordingly, the trade and transportation margins are not distributed. Second, the IO framework reports taxes net subsides. Third, due to statistical differences, the taxes net subsidies reported in the IO model do not always exactly match the taxes minus subsides on products from the Government Accounts.

Taking these issues into consideration, the 'Taxes on products, receivable' are modelled as follows. First, the taxes net subsides on final and intermediate products are calculated as:

$$Net\_Taxes\_products_{riut} = Exo\_Tax\_rate\_products\_domestic_{riut} \times \Big[Final\_demand\_domestic\_BP\_real_{riut} \times Price_{rit} \times (1 + Exo\_Margins\_paid\_domestic_{riut}) - Exo\_Margins\_rec\_domestic_{riu} \times \sum_j \Big(Final\_demand\_domestic\_BP\_real_{rjut} \times Price_{rjt} \times \big(1 + Exo\_Margins\_paid\_domestic_{rju}\big)\Big)\Big] + \sum_s \Big\{Exo\_Tax\_rate\_products\_import_{sriut} \times \Big[Final\_demand\_import\_BP\_real_{sriut} \times Price_{sit} \times (1 + Exo\_Margins\_paid\_import_{sriu}) - Exo\_Margins\_rec\_import_{sriu} \times \sum_j \Big(Final\_demand\_import\_BP\_real_{srjut} \times Price_{sjt} \times \big(1 + Exo\_Margins\_paid\_import_{srju}\big)\Big)\Big]\Big\}$$

Equation 41

As already mentioned, there are some statistical differences between the data in the IO and the data in the Government Accounts. This statistical difference is included in the model to keep the consistency between the two accounting frameworks. The term $Taxes\_products\_Stat\_diff_{rt}$ is an exogenous parameter that represents the difference between the total taxes net subsides on products reported in the Government accounts and the ones reported in the IO tables.

Then, it follows that the taxes net subsides on products adjusted to Government accounts can be calculated as:

$$Net\_Taxes\_products\_Gov\_accounts_{rt} = \sum_{iu} Net\_Taxes\_products_{riut} + Exo\_Net\_Taxes\_products\_Stat\_diff_{rt}$$

Equation 42

Finally, taxes net subsides are split up, following exogenous structures,

$$Taxes\_products_{rt} = Net\_Taxes\_products\_Gov\_accounts_{rt} \times Exo\_Structure\_Net\_Taxes\_products\_Gov\_accounts_{r,taxes,t}$$

Equation 43

The next component of the government revenue is the 'Other taxes and subsidies on production, receivable'. This component is calculated multiplying the gross output by country and industry from the IO model with the exogenous tax rate:

$$Net\_Taxes\_output_{rit} = \sum_i Exo\_Tax\_rate\_output_{rit} \times Output_{rit}$$

Equation 44

Similarly, taxes net subsides on production are calculated as:

$$Taxes\_output_{rit} = Net\_Taxes\_output_{rit} \times Exo\_Structure\_Net\_Taxes\_output_{r,taxes,t}$$

Equation 45

The government property income is the product of the total stock of assets of the government and an exogenous interest rate:

$$Gov\_Property\_income_{rit} = Exo\_Rate\_Gov\_Property\_income_{rit} \times Gov\_Assets_{rit-1}$$

Equation 46

Next, 'Current taxes on income, wealth etc.' is calculated as the sum of the taxes on the lagged operating surplus from the production submodule, income received by each household (h) and their stock of assets:

$$Tax\_Income\_Wealth_{rit} = \sum_h Number\_households_{rht} \times \{\sum_i [Share\_Labour\_income_{rhit} \times (1 - Exo\_rate\_Social\_security_{rit}) \times Labour\_compensation_{rit} + Share\_Operating\_surplus_{rhit} \times Exo\_Tax\_rate\_Operating\_Surplus_{rit} \times Gross\_Operating\_surplus_{rit}] +$$

$Exo\_Tax\_rate\_Income_{rht} \times HH\_Income_{rht-1} + Exo\_Tax\_rate\_Wealth_{rht} \times Net\_wealth_{rht-1}\}$

Equation 47

Net social contributions are calculated as the sum of employers' and households' social contribution. Employers' social contributions are obtained by multiplying the (gross) 'Labour compensation' by industry from the labour submodule times the exogenous rate of social contributions:

$$Social\_security\_contributions_{rt} = \sum_{hi} Number\_households_{rht} \times Share\_Labour\_income_{rhit} \times (1 - Exo\_rate\_Social\_security_{rit}) \times Labour\_compensation_{rit}$$

Equation 48

Finally, 'Other Government revenue' is calculated as a fixed share of total value added:

$$Gov\_Other\_revenue_{rt} = Exo\_rate\_Gov\_Other\_revenue_{rt} \times \sum_{i} Value\_added_{rit}$$

Equation 49

The total government revenue is the sum of the components of the Government Accounts:

$$Gov\_revenue_{rt} = Taxes\_products_{rt} + Taxes\_output_{rt} + Gov\_Property\_income_{rt} + Taxes\_Income\_Wealth_{rt} + Social\_security\_contributions_{rt} + Gov\_Other\_revenue_{rt}$$

Equation 50

Now, we can calculate the ratio of government revenue to GDP, which will be used to calculate the error in the next period as for:

$$Gov\_Revenue\_to\_GDP_{\ rt} = Gov\_Revenue_{rt} / GDP_{rt}$$

Equation 51

4.4.3.2.2. GOVERNMENT EXPENDITURE

The equations for the components of the Government expenditure are depicted below.

Subsidies on products and production are calculated as:

$$Subsidies\_products_{rt} = Net\_Taxes\_products\_Gov\_accounts_{rt} \times Exo\_Structure\_Net\_Taxes\_products\_Gov\_accounts_{r,subsidies,t}$$

Equation 52

$$Subidies\_output_{rt} = Net\_Taxes\_output_{rit} \times Exo\_Structure\_Net\_Taxes\_output_{r,subsidies,t}$$

Equation 53

Total interest paid by the government is calculated by multiplying the total stock of debt at the beginning of the time step with the interest rate paid for debt:

$$Debt\_Interest_{rt} = Exo\_Debt\_Interest\_rate_{rt} \times Gov\_Debt_{rt-1}$$

Equation 54

The actual value of the total social benefits at the end of the period is calculated by multiplying the total number of retired people (from the socio-demographic module) and the unemployed people (product of active population from the socio-demographic module and unemployment rate) with the average (exogenous) annual benefit per person (exogenous):

$$Social\_benefits_{rt} = Retired_{rt} \times Exo\_Pension_{rt} + Active\_population_{rt} \times Unemployment\_rate_{rt} \times Exo\_Unemployment\_benefit_{rt}$$

Equation 55

The other expenditure categories (Final consumption expenditure; Gross fixed capital formation; Other current transfers, payable; and Other expenditure) are calculated following exogenous structures over total expected Expenditure:

$$Gov\_Consumption_{rt} = Gov\_Expenditure\_obj_{rt} \times Exo\_Structure\_Gov\_Expenditure_{r,GovCons,t}$$

Equation 56

$$Gov\_Investment_{rt} = Gov\_Expenditure\_obj_{rt} \times Exo\_Structure\_Gov\_Expenditure_{r,GovGFCF,t}$$

Equation 57

$$Gov\_Transfers_{rt} = Gov\_Expenditure\_obj_{rt} \times Exo\_Structure\_Gov\_Expenditure_{r,GovTransf,t}$$

Equation 58

$$Gov\_Other\_Expenditure_{rt} = Gov\_Expenditure\_obj_{rt} \times Exo\_Structure\_Gov\_Expenditure_{r,OtherExp,t}$$

Equation 59

The government final consumption expenditure and gross fixed capital formation are split up by function according to the exogenous COFOG structures:

$$Gov\_Consumption\_COFOG_{rgt} = Exo\_Structure\_Gov\_Consumption\_COFOG_{rgt} \times Gov\_Consumption_{rt}$$

Equation 60

$$Gov\_Investment\_COFOG_{rgt} = Exo\_Structure\_Gov\_Investment\_COFOG_{rgt} \times Gov\_Investment_{rt}$$

Equation 61

Where the subscript g denotes the COFOG categories.

These two variables are transformed into final demand by industry (in purchaser prices) using exogenous bridge matrices matrix:

$$Final\_demand\_PP_{ri,GGFC} = Exo\_Structure\_Gov\_Consumption\_Industry_{rijt} \times Gov\_Consumption\_COFOG_{rgt}$$

Equation 62

$$Gov\_Final\_demand\_PP_{ri,GFCF} = Exo\_Structure\_Gov\_Investment\_Industry_{rijt} \times Gov\_Investment\_COFOG_{rgt}$$

Equation 63

Note that government investment is part of the vector of total investment ($Final\_demand\_PP_{ri,GFCF}$), which includes both private and public investment. The last two equations will be further transformed into vectors of final demand in real terms and basic prices.

Finally, the actual (i.e., not projected) value of total government expenditure at the end of the period is the sum of the components of the government accounts calculated as:

$$Gov\_Expenditure_{rt} = Debt\_Interest_{rt} + Social\_benefits_{rt} + Subsidies\_Products_{rt} + Subsidies\_Production_{rt} + Gov\_Consumption_{rt} + Gov\_Investment_{rt} + Gov\_Transferences_{rt} + Gov\_Other\_Expenditure_{rt}$$

Equation 64

4.4.3.2.3. TOTAL GOVERNMENT REVENUE, EXPENDITURE AND BALANCE

The difference between the total government revenue and the total government expenditure gives the actual government deficit/surplus at the end of the period (i.e., 'Net lending (+)/Net borrowing (-)')

$$Gov\_Deficit\_Surplus_{rt} = Gov\_Revenue_{rt} - Gov\_Expenditure_{rt}$$

Equation 65

The ratio government deficit/surplus to GDP, used to calculate the error in the next period, is

$$Gov\_Deficit\_Surplus\_to\_GDP_{rt} = Gov\_Deficit\_Surplus_{rt}/GDP_{rt}$$

Equation 66

The Government balance is transformed into the change in the stock of gross debt by adding 'Net acquisition (+) of financial assets', 'Total adjustments', and 'Statistical difference'. The 'Net acquisition (+) of financial assets' variable comes from the financial module, while the 'Total adjustments' and 'Statistical difference' are set exogenously:

$$Change\_Gov\_Debt_{rt} = -Gov\_Deficit\_Surplus_{rt_{rt}} + Gov\_Assets\_Net\_Adquisition_{rt} + Exo\_Gov\_Debt\_Adjustment_{rt} + Exo\_Gov\_Debt\_Stat\_Diff_{rt}$$

Equation 67

Finally, the resulting stock of government debt at the end of the period is calculated as:

$$Gov\_Debt_{rt} = Gov\_Debt_{rt-1} + Change\_Gov\_Debt_{t-q}$$

Equation 68

This variable will be used in the next period to calculate interest payments.

4.4.3.3. THE PRODUCTION/FIRMS SUBMODULE

4.4.3.3.1. THE INPUT-OUTPUT QUANTITY MODEL

The starting point of the IO quantity model are the equations describing the IO loop between production and demand in the quantity model:

$$Output\_real_{rkt} = \sum_j Intermediates\_domestic\_real_{rkjt} + \sum_c Final\_demand\_domestic\_BP\_real_{rkct} + \sum_s Exports_{rskt}$$

Equation 69

$$Intermediates\_real_{rkjt} = Exo\_Technical\_coefficients_{rkjt} \times Output\_real_{rjt}$$

Equation 70

These equations are solved using the Leontief Inverse matrix.

On the one hand, the IO model computes the real output (i.e., output in constant prices of 2015) of non - energy industry k of region r in year t as the sum of the domestic intermediate and final demand of its products plus its exports. On the other hand, it computes the total (i.e., domestic plus imported) intermediate inputs produced by the non-energy industry k required by industry j of region r to satisfy its demand being $Exo\_A\_matrix_{rijt}$ the matrix of technical coefficients whose element a_{rijt} represents the inputs of products from industry i required by industry j of region r to produce one unit of output. The elements of this matrix are exogenous except for the case of the demand of products from energy-related industries, for which these coefficients are endogenously calculated from the link with the bottom-up with the energy module. Note that these equations do not compute the output of energy industries and the intermediate demand of energy inputs. The calculation of these variable is described in section 4.4.3.3.7, where we show the link with the bottom-up energy module.

4.4.3.3.2. INTERNATIONAL TRADE SUBMODULE

The variable $Final\_demand\_domestic\_BP\_real_{rict}$, represents the domestic final demand final demand at basic prices (BP) in real terms by final demand category c, and it is calculated as[6]

$$Final\_demand\_domestic\_BP\_real_{rict} = (1 - Import\_shares_{rict}) \times Final\_demand\_BP\_real_{rict}$$

Equation 71

Where $Final\_demand\_BP\_real_{ict}$ is the total (i.e., domestic and imported) final demand at basic prices (BP) in real terms, which components are:

- $Final\_demand\_BP\_real_{ri,HFCE,t}$ household consumption, is a function of income, wealth and prices.
- $Final\_deman\_BP\_real_{ri,NPISH,t}$ consumptions of Non-profit institutions serving households prices (exogenous)
- $Final\_demand\_BP\_real_{ri,GGFC,t}$ government consumption $Final\_demand\_BP\_real_{ri,GFCF,t}$ gross fixed capital formation (endogenous) (see section 4.4.3.3.5).
- $Final\_demand\_BP\_real_{ri,INVNT,t}$ Change in inventories and valuables (exogenous).

The variable $Import\_shares_{rict}$ represents the share of the final demand that is satisfied with imported goods which is derived from the Armington elasticities calculated for final demand and intermediate products (see Deliverable 4.5) and is calculated as:

$$log\left[\frac{Import\_shares_{rict}}{(1 - Import\_shares_{rict})}\right] = Constant\_Final\_import_{rict} + Beta\_Final\_import\_1_i \times log(Import\_shares_{rict-1}) + Beta\_Final\_Import\_shares\_2_i \times log\left[\frac{Price_{rict-1}}{Price\_Imports_{rict-1}}\right]$$

Equation 72

The domestic intermediates are calculated analogously to the final demand.

The imported final and intermediate demand is defined as

$$Final\_demand\_import\_BP\_real_{srict} = Import\_origin\_shares_{srit} \times Import\_shares_{rict} \times Final\_demand\_BPs\_real_{rict}$$

Equation 73

$$Intermediates\_import\_real_{srijt} = Import\_origin\_shares_{srit} \times Import\_shares_{rijt} \times Intermediates\_real_{rijt}$$

Equation 74

Where $Import\_origin\_shares_{srit}$ origin shares of trade the trade structure by origin of the base year.

Finally, the exports of each region r are calculated as the sum of the imports of all the other regions s from region r:

[6] Annex 1 of Deliverable 4.6 shows the equations for the transformation of the final demand from real to nominal terms and from basic to purchaser's prices.

$$Exports\_real_{rsit} = \sum_c Final\_demand\_import\_BP\_real_{rsict} + \sum_j Intermediates\_import\_real_{rsijt}$$

Equation 75

4.4.3.3.3. LABOUR SUBMODULE

According to the concepts applied in national accounts, this primary distribution of income covers the components of labour compensation (which in WILIAM includes compensation of employees and mixed income), capital compensation and taxes less subsidies on products.

Labour demand in number of hours is calculated on the basis of the relationship between real output and labour productivity by industry:

$$Hours_{rit} = \frac{Output\_real_{rit}}{Labour\_productivity_{rit} \times (1 - L\_damage\_CC\_rate(T)_{rit})}$$

Equation 76

Where $L\_damage\_CC\_rate(T)_{rit}$ reflects the reduction in labour productivity due to climate change (see Deliverable 4.7) and $Labour\_productivity_{rit}$ is the labour productivity (exogenous)..

The labour compensation in nominal terms is then calculated by multiplying the wage per hour by the number of hours:

$$Labour\_Compensation\_nominal_{rit} = Wage\_per\_hour_{rit} \times Hours_{rit}$$

Equation 77

The wage equation defines for each industry i of each region r the wages (per hour) as a function of the change in the aggregate consumer price PCP_{rt}, the change in the industry-level hour productivity $Labour\_productivity_{rit}$, and the actual unemployment rate in relation to full employment rate:

$$Wage\_per\_hour_{rit} = Wage\_per\_hour_{rit-1} \times \left[1 + \exp\left(\alpha_{ri} + \beta_{1i} log\left(\frac{Labour\_productivity_{rit}}{Labour\_productivity_{rit-1}}\right) + \beta_{2i} log\left(\frac{PCP_{rt-1}}{PCP_{rt-2}}\right) + \beta_{3i} log\left(\frac{Unemployment\_rate_{rt-1}}{Full\_Unemployment\_rate_{rt-1}}\right)\right)\right]$$

Equation 78

The average wage per hours is:

$$Wage\_per\_hour\_average_{rt} = \frac{\sum_i Labour\_Compensation\_nominal_{rit}}{\sum_i Hours_{rit}}$$

Equation 79

Employment by industry is caudated as the quotient between the total hours worked and the average hours worked per worker (exogenous):

$$Employment_{rit} = \frac{Hours_{rit}}{Exo\_Hours\_worker_{rit}}$$

Equation 80

Total employment is calculated summing the employment over all industries:

$$Employment\_total_{rt} = \sum_i Employment_{rit}$$

Equation 81

On the other hand, labour supply is a function of the working age population from the demographic module) and the participation rate:

$$log(Particiaption\_rate_{rt}) = Constant\_Particiaption\_rate_{rt} + Beta\_Particiaption\_rate\_1_{rt} \times log(1 - Unemployment\_rate_{rt-1}) + Beta\_Particiaption\_rate\_2_{rt} \times log\left[\frac{Wage\_per\_hour\_average_{rt-1}}{Consumer\_Price\_Index_{rt-1}}\right]$$

Equation 82

$$Labour\_force_{rt} = Participation\_rate_{rt} \times Working\_age\_population_{rt}$$

Equation 83

Finally, unemployment rate is calculated as:

$$Unemployment\_rate_{rt-1} = 1 - \frac{Employment\_total_{rt}}{Labour\_force_{rt}}$$

Equation 84

4.4.3.3.4. VALUE ADDED, GROSS DOMESTIC PRODUT AND NET DOMESTIC PRODUCT

The value added by industry in nominal terms is calculated as the difference between the output minus the intermediate demand and the net taxes on intermediate products:

$$Value\_added_{rit} = Price_{rit} \times Output\_real_{rit} - \sum_j Price_{rjt} \times Intermediates\_domestic\_real_{rjit} - \sum_{sj} Price_{sjt} \times Intermediates\_import\_real_{srjit} - \sum_i Net\_Taxes\_products_{rijt}$$

Equation 85

The gross operating surplus is calculated by subtracting the taxes on production and labour compensation from value added:

$$Gross\_Operating\_surplus\_nominal_{rit} = Value\_added_{rit} - (Labor\_Compensation\_nominal_{rit} + Net\_Taxes\_output_{rit})$$

Equation 86

Finally, we can compute nominal and real Gross Domestic Product (GDP) as:

$$GDP_{rt} = \sum_i [Price_{rit} \times Output\_real_{rit} - \sum_j Price_{rjt} \times Intermediates\_domestic\_real_{rjit} - \sum_{sj} Price_{sjt} \times Intermediates\_import\_real_{srjit} + \sum_c Net\_Taxes\_products_{rict}]$$

Equation 87

The GDP in real terms is calculated as:

$$GDP\_real_{rt} = \sum_i [Output\_real_{rit} - \sum_j Intermediates\_real_{rijt} + \sum_c Net\_Taxes\_products\_real_{rict}]$$

Equation 88

Where $Net\_Taxes\_products\_real_{rict}$ are the net taxes on products in real terms which are calculated using the tax rates of the base year ($t = 0$):

$$Net\_Taxes\_products\_real_{rict} = Exo\_Tax\_rate\_products\_domestic_{ric0} \times \left[Final\_demand\_domestic\_BP\_real_{rict} \times (1 + Exo\_Margins\_paid\_domestic_{rict}) - Exo\_Margins\_rec\_domestic_{ric} \times \sum_j \left(Final\_demand\_domestic\_BP\_real_{rjct} \times (1 + Exo\_Margins\_paid\_domestic_{rjc})\right)\right] + \sum_s \left\{Exo\_Tax\_rate\_products\_import_{sric0} \times \left[Final\_demand\_import\_BP\_real_{srict} \times (1 + Exo\_Margins\_paid\_import_{sric}) - Exo\_Margins\_rec\_import_{sric} \times \sum_j \left(Final\_demand\_import\_BP\_real_{srjct} \times (1 + Exo\_Margins\_paid\_import_{srjc})\right)\right]\right\}$$

Equation 89

In addition, by subtracting the consumption of fixed capital from the real GDP, we obtain the Net Domestic Product in real terms:

$$NDP\_real_{rit} = GDP\_real_{rit} - Consumption\_fixed\_capital\_real_{rit}$$

Equation 90

In the context of damages caused by climate change and the impact of mitigation and adaptation policies, this variable can be seen as an alternative to GDP for measuring the aggregate economic impact of different change scenarios, since it considers the losses in the stock of capital due to climate change and, therefore, comprises some wealth and welfare impacts.

4.4.3.3.5. GROSS FIXED CAPITAL FORMATION AND CAPITAL STOCK

This section introduces the new specification for the gross fixed capital formation (GFCF) or investment (for simplification) and the stock of capital. In particular, the WILIAN model goes a step beyond MEDEAS by 1) improving the calculation of the GFCF by industry[7], 2) linking this variable to the vector of GFCF by product of the final demand of the IO framework, and 3) introducing explicitly the stock of capital as part of the production process.

The starting point of the calculation of GFCF is the capital stock it needs (long-term optimal stock):

$$Desired\_Capital\_stock\_real_{rit} = \frac{Output\_real_{rit}}{Capital\_productivity_{rit}}$$

Equation 91

The equation of (actual) capital accumulation in real terms is explicitly defined as:

$$\Delta Capital\_Stock\_real_{rit} = GFCFt\_PP\_real_{rit-1} - Consumption\_fixed\_capital\_real_{rit-1}$$

Equation 92

where $GFCF\_PP\_real_{rit}$ is the real gross fixed capital formation in purchaser's prices and $Consumption\_fixed\_capital\_real_{rit-1}$ is the consumption of fixed capital in the previous period calculated ass:

$$Consumption\_fixed\_capital\_real_{rit} = Exo\_Capital\_depreciation\_rate_{ri} \times Capital\_stock\_real_{rit} + Capital\_damage\_CC\_rate(T)_{rit} \times Capital\_stock\_real_{rit} = Capital\_Depreciation_{rit} + Capital\_Damage\_CC_{rit}$$

Equation 93

Damage caused by climate change at first reduces the capital stock and can be offset in the next period through capital formation. Therefore, the investment behaviour in WILIAM is specified as in 'stock adjustment' models. These 'stock adjustment' models are dynamically specified for the capital stock, i.e., $\Delta Capital\_stock\_real_{rit}$ is a function of a 'stock adjustment' term: $Exo\_Capital\_stock\_adjustment \times [Capital\_stock\_real^{*}_{ri\,t} - logCapital\_stock\_real_{rit}]$ with $Exo\_Capital\_stock\_adjustment > 0$. This term works similar to an 'error correction' model. Since $Capital\_stock_{rit}$ includes the effect of capital consumption due to damages $Capital\_Damage\_CC_{it}$, the term of 'stock adjustment' also determines to what extent the investment compensates these capital stock losses caused by damages. Therefore, the parameter $Exo\_Capital\_stock\_adjustment$ captures both the effects of low capital utilization and

[7] In principle, the link between GFCF by industry and the vector of GFCF was intended to be solved by estimating the sectoral GFCF independently from the IO framework, undertaking an analysis based on the microdata provided by the database ORBIS. This proposal has not been taken into account in the end, because a better option was found in the course of the development of the model. This option not only permits to know the GFCF by sector, but also the composition of this variable in terms of types of investment goods.

damage from climate change. Thus, we can specify an investment equation that contains the 'stock adjustment':

$$GFCF\_PP\_real_{rit} = Exo\_Capital\_stock\_adjustment_i \times [Capital\_stock\_real^*_{ri\,t} - Capital\_stock\_real_{rit}]$$

Equation 94

Note that for some energy sectors the GFCF is calculated from the bottom-up physical models. In particular, the investment in the different power producing industries is calculated using data from the capacity expansion (MW) computed in the energy submodule, which is transformed into monetary terms. Also, in the case of government GFCF is calculated in the government submodule.

The real GFCF by industry calculated is transformed into a vector of GFCF by type of product using a set of transformation matrices estimated using data from EUROSTAT and EUKLEMS. These matrices show the structure (i.e., type of investment goods) of the GFCF of each investing industry (excluding Government investment). The real GFCF by type of investment good in purchaser's prices is:

$$GFCF\_by\_good\_PP\_real_{rjt} = \sum_i GFCF\_PP\_real_{rit} \times Exo\_GFCF\_structure_{rijt}$$

Equation 95

These vectors of investment will be further transformed into a vector of final demand in basic prices $Final\_demand\_BP\_real_{ri,GFCF,t}$ (see section 4.4.3.3.6) and linked to the IO model.

The price of GFCF of each (investing) industry is calculated as the product of the deflator at purchaser's prices (see Annex 1) and the matrix of investment structure:

$$Price\_GFCF_{rit} = \sum_j Deflator\_Final\_demand\_PP_{rj,GFCF,t-1} \times Exo\_GFCF\_structure_{rji}$$

Equation 96

Finally, the capital stock is calculated as follows:

$$K\_real_{rit} = GFCF\_PP\_real_{rit} + K_{rit-1} \times [1 - Exo\_capital\_depreciation\_rate_{ri} - K\_damage\_CC\_rate(T)_{rit}]$$

Equation 97

4.4.3.3.6. THE INPUT-OUTPUT PRICE MODEL AND VALUATION (NOMINAL/REAL TERMS, BASIC/PRUCHASER PRICES)

The price model includes a mark-up and the costs of capital, labour, intermediate consumption (domestic and imported), and taxes. The capital cost contains the consumption of fixed capital per unit of output. This is defined by the two capital consumption rates (depreciation and damages) and the industry capital formation price index ($Price\_GFCF_{ri}$) defined through the investment matrix. For the case of labour, the price equation also reflects the effect of CC on productivity ($L\_damage\_CC\_rate(T)_{rit}$). The general expression of the price equation[8] is:

$$Price_{rit} = \Big[Exo\_Mark\_up_{ri} + Price\_GFCF_{rit} \times (depreciation\_rate_{ri} + K\_damage\_CC\_rate(T)_{rit}) \times \frac{1}{K\_productivity_{ri\,t-1}} + \frac{Wage\_per\_hour_{rit}}{Wage\_per\_hour_{ri0}} \times \frac{1}{Exo\_Labour\_productivity_{rit} \times L\_damage\_CC\_rate(T)_{rit}} + \sum_{sj} Price_{rjt} \times (1 - Import\_shares_{srjit}) \times Exo\_Technical\_coeffcients_{rjit} + \sum_{sj} Price_{sjt} \times Import\_shares_{srjit} \times Exo\_Technical\_coeffcients_{rjit} + Exo\_Tax\_rate\_Output_{rit} \times Price_{rit} + \frac{\sum_j Net\_Taxes\_products_{rjit}}{Output\_real_{rit}}\Big]$$

Equation 98

[8] This is a simultaneous equation since the price is included in both the left- and right-hand sides. Thus, it has been adapted to be implemented in Vensim®.

This equation is used to calculate the output prices of all the sectors in WILIAM except the sectors related to material extraction, for which prices are calculated in the materials module.

Prices are also used in the model to deflate variables (i.e., to transform variables expressed in nominal terms to real terms and vice versa). In addition, in the model two different valuation methods for final goods coexist:

- The basic price is the amount receivable by the producer from the purchaser for a unit of a good or service produced as output minus any tax payable, and plus any subsidy receivable, by the producer as a consequence of its production or sale. It excludes any transport charges invoiced separately by the producer.
- The purchaser price is the amount paid by the purchaser, excluding any value added tax or similar tax deductible by the purchaser, in order to take delivery of a unit of a good or service at the time and place required by the purchaser. The purchaser price of goods includes any transport charges paid separately by the purchaser to take delivery at the required time and place.

4.4.3.3.7. LINK ECONOMY – ENERGY

This section describes the linkages between the economy and energy modules. Similar linkages are currently under development for the materials and land modules.

As already pointed in section 4.4.3.3.1, while non-energy input coefficients are exogenous, the energy-related intermediate consumption are endogenous and depend on the evolution of different variables calculated in the energy module. As a result, the matrix of technical coefficients is an hybrid between static coefficients for the demand of non-energy inputs and dynamic coefficients of the demand of energy inputs. This approach is particularly relevant for the overall consistency of the whole model, since it guarantees that changes in the energy sphere are reflected in the economic sphere and vice versa.

In order to implement this approach, the 62 industries of the WILIAM sector have been separated into two different groups (see section 4.4.3.3.1): 43 non-energy-related sectors ($\boldsymbol{k}$) and 19 energy-related industries (see D4.2). The latter are classified into three categories: energy-extracting industries and industries supplying energy for transformation ($\boldsymbol{e}$), industries transforming energy inputs into final energy ($\boldsymbol{q}$) and by households ($\boldsymbol{f}$).

The link between the economic and energy modules can be summarised in two steps (the chart in Annex 5, provides a simple graphical presentation of these linkages):

1) The economic module computes the demand of final energy using energy intensities coming from the energy module. This demand of final energy is sent to the energy module.
2) The energy module decides how the final energy is produced. In particular, the energy module calculates the transformation output of each energy-related industry, the inputs for transformation the primary energy production, and the own energy consumption of the energy industries. This information is passed to the economic module and linked to the different parts of the Economy module.

4.4.3.3.7.1. FROM ECONOMY TO ENERGY

Starting by the first stage, we differentiate two components in the demand of final energy: the intermediate demand by industries and the final demand by components of the final demand. This section only covers the demand of final energy for intermediate uses. The demand of final energy by the final demand (i.e., households) is described in the consumption submodule.

The physical intermediate demand of final energy (in EJ) delivered by energy industry f to non-energy industry k is calculated as the product of the (monetary) output in real terms times the energy intensity:

$$Intermediates\_Phy_{rfkt} = Final\_energy\_intensity_{rfkt} \times Output\_real_{rkt}$$

Equation 99

The $Final\_energy\_intensity_{rfkt}$ is the energy demand of final energy delivered by industry f (e.g., mining of coal and lignite; extraction of peat) per unit of output of non-energy industry k. The products of the industries selling final energy are linked to the final energy categories reported in the energy balances.

The final energy intensities of the different sectors and energy products ($Final\_energy\_intensity_{rfkt}$) are modelled in the energy module and can be computed from a bottom-up approach (for those sectors with modelled end-use technological disaggregation) or from a top-down estimation (for the remaining sectors).

The physical intermediate demand of final energy only includes the demand of non-energy sectors (it does not include the energy consumption of energy sectors: e.g., diesel in mining, electricity in power stations, self-consumption of oil distillates in refineries, etc.) and transformation inputs[9] (e.g., coal in power plants, crude oil in refineries, etc.)).

On the other hand, the monetary intermediate demand of final energy delivered by energy industry f to non-energy industry k is calculated as the product of the physical intermediate demand and the implicit price:

$$Intermediates\_real_{rfkt} = Intermediates\_Phy_{rfkt} \times Price\_implicit\_2015_{rfk}$$

Equation 100

$Price\_implicit\_2015_{rfk}$ are the implicit prices and is the ratio between the monetary and physical final energy demand in the base year (2015) and remain constant in the simulations. The implicit prices work as a conversion factor in the model and are used to convert monetary units, in real terms, to physical units and vice versa. Contrarily to implicit prices, market prices change during the simulations and respond to different factors such as scarcity, changes in production costs, taxes, etc.

Finally, the monetary output of the industries selling final energy is calculated as the sum of the monetary intermediate demand of final energy, the domestic final demand from the household submodule and the exports:

$$Output\_real_{rft} = \sum_i Intermediates\_domestic\_real_{rfit} + \sum_c Final\_demand\_domestic\_BP\_real_{rfct} + \sum_{s,i} Exports_{rsfit}$$

Equation 101

The monetary output of the industries selling final energy is linked to the total output ($Output\_real_{rjt}$) and calculates the non-energy inputs required by the industries delivering final energy, given the final energy intensities calculated in the energy module.

Note this procedure method allows endogenizing the final energy-related technical coefficients given that they are defined as $Intermediate\_demand\_Mon_{rfkt}/Output\_Mon_k$.

4.4.3.3.7.2. FROM ENERGY TO ECONOMY

[9] Following the terminology from the Eurostat Energy Balances: https://ec.europa.eu/eurostat/documents/38154/4956218/ENERGY-BALANCE-GUIDE-DRAFT-31JANUARY2019.pdf/cf121393-919f-4b84-9059-cdf0f69ec045.

In the second step, the energy module calculates the transformation inputs ($Input\_Transf_{rnpt}$) by type of energy (n: transformation input, p: transformation process) required to satisfy the final energy demand coming from the economic module. This information is passed to the economic module as follows.

First, the physical inputs by energy type and transformation process coming from the energy module has to be converted into the industry classification of the economic module. This conversion results in the physical intermediate demand of energy by the industries transforming energy inputs into final energy:

$$Intermediate\_demand\_Phy_{reqt} = \sum_{n,p} Input\_Transf_{rnpt} \times EXO\_Correspondence\_Input_{rnpeqt}$$

Equation 102

Where $Correspondence\_Input_{rnpeqt}$ are correspondence matrices (i.e., matrices with ones and zeroes) of appropriate dimension used to convert the energy flows from the classification of the energy module to the classification of the economic module.

Similarly, the physical own consumption of final energy by type and transformation process coming from the energy module has to be converted into the industry classification of the economic module. This conversion results in the physical intermediate demand of final energy by the industries transforming energy inputs into final energy:

$$Intermediates\_Phy_{rfqt} = \sum_{n}, pConsumption\_Energy\_Sector_{rnpt} \times EXO\_Correspondence\_Input_{rnpfqt}$$

Equation 103

The physical intermediate demand of the industries transforming energy inputs into final energy is converted into monetary units using implicit prices:

$$Intermediates\_real_{reqt} = Intermediated\_Phy_{reqt} \times Price\_implicit\_2015_{req}$$

Equation 104

$$Intermediates\_real_{rfqt} = Intermediates\_Phy_{rfqt} \times Price\_implicit\_2015_{rfq}$$

Equation 105

In the above equation, q is the industry which demands (in this case transformation), e is the industry which supplies the energy to be transformed and f is the industry supply final energy for the own consumption of the energy transformation sector.

Again, this method allows endogenizing the energy-related technical coefficients given that they are defined as $Intermediates\_real_{reqt}/Output\_real_{rqt}$ and $Intermediates\_real_{rfqt}/Output\_real_{rqt}$

Finally, the output to the energy-extracting industries is the sum of the domestic intermediate demand of the industries transforming energy inputs into final energy, plus the direct sales of final energy products of the energy-extracting industries plus exports:

$$Output\_real_{ret} = \sum_{k} Intermediates\_domestic\_real_{reqt} + \sum_{k} Intermediates\_domestic\_real_{rekt} + \sum_{c} Final\_demand\_domestic\_BP\_real_{rect} + \sum_{sj} Exports_{rsejt}$$

Equation 106

The monetary output of the energy-extracting industries is linked to the total output ($Output\_real_{rjt}$).

The physical own consumption of final energy-extracting industries is calculated as

$$Intermediates\_Phy_{rfet} = \sum_{n} Consumption\_Energy\_Sector_{rnht} \times EXO\_Correspondence\_Input_{rnpfet}$$

Equation 107

The physical intermediate demand of the energy-extracting industries is converted into monetary units using implicit prices:

$$Intermediates\_real_{rfet} = Intermediates\_Phy_{rfet} \times Price\_implicit\_2015_{rfe}$$

Equation 108

4.4.4. DATA SOURCES

In this subsection specific data sources for calibration and validation processes of the module will be presented.

4.4.4.1. DATA FOR THE HOUSEHOLD/ CONSUMPTION SUBMODULE

For EU countries:

- Micro-data surveys:
  - European Union Statistics on Income and Living Conditions (SILC)[10] from Eurostat (https://ec.europa.eu/eurostat/web/microdata/european-union-statistics-on-income-and-living-conditions);
  - Household Budget Survey (HBS)[11] from Eurostat (https://ec.europa.eu/eurostat/web/household-budget-surveys);
  - Household Finance and Consumption Survey (HFCS)[12] from European Central Bank (https://www.ecb.europa.eu/pub/economic-research/research-networks/html/researcher_hfcn.en.html).
- National Accounts:
  - Non-financial transactions from Eurostat (nasa_10_nf_tr).
  - Final consumption expenditure of households by consumption purpose (COICOP 3 digit) from Eurostat (nama_10_co3_p3).
- Other secondary sources:
  - Financial balance sheets from Eurostat (nasa_10_f_bs).
  - Balance sheets for non-financial assets from Eurostat (nama_10_nfa_bs).
  - Population change - Demographic balance and crude rates at national level from Eurostat (demo_gind).
  - HICP (2015 = 100) - annual data (average index and rate of change) from Eurostat (prc_hicp_aind).
  - Gross efficiency heat and gross efficiency transport index from Odyssee database (https://www.indicators.odyssee-mure.eu/energy-efficiency-database.html)
  - Prices of natural gas and electricity (2015) from Eurostat (nrg_pc_h).
  - Prices of fuels (2015) from Weekly Oil Bulletin (https://ec.europa.eu/energy/data-analysis/weekly-oil-bulletin_enhttps://ec.europa.eu/eurostat/web/energy/data/database).

10 SILC contains confidential data granted by Eurostat for scientific purposes. For more information on the access to confidential data, see the Commission Regulation (EU) No 557/2013 of 17 June 2013 implementing Regulation (EC) No 223/2009 of the European Parliament and of the Council on European Statistics as regards access to confidential data for scientific purposes and repealing Commission Regulation (EC) No 831/2002.

11 HBS contains confidential data granted by Eurostat for scientific purposes. For more information on the access to confidential data, see the Commission Regulation (EU) No 557/2013 of 17 June 2013 implementing Regulation (EC) No 223/2009 of the European Parliament and of the Council on European Statistics as regards access to confidential data for scientific purposes and repealing Commission Regulation (EC) No 831/2002.

12 HFCS contains confidential data granted by the European Central Bank for scientific purposes only and under a confidentiality commitment.

- Passenger kilometres per mode and country: Global Transportation Demand Development with Impacts on the Energy Demand and Greenhouse Gas Emissions in a Climate-Constrained World, School of Energy Systems, LUT University, Finland, https://doi.org/10.3390/en12203870 (Khalili et al., 2019)

For non-EU countries:

- National Accounts:
  - Final consumption expenditure of households from OECD (https://stats.oecd.org/Index.aspx?DataSetCode=SNA_TABLE5).
  - 14A. Non-financial accounts by sectors from OECD (https://stats.oecd.org/Index.aspx?DataSetCode=SNA_TABLE14A).
- Other secondary sources:
  - Households' financial assets and liabilities from OECD (https://stats.oecd.org/Index.aspx?DataSetCode=QASA_7HH).
  - Households and NPISHs final consumption expenditure from World Bank (https://data.worldbank.org/indicator/NE.CON.PRVT.ZS).
  - Global Consumption Database from World Bank (https://datatopics.worldbank.org/consumption/home).

4.4.4.2. DATA FOR THE GOVERNMENT SUBMODULE

Regarding the Government submodule, the main data sources for the parametrisation in the government submodule are Eurostat, OECD and IMF:

- Annual government finance statistics (code gov_10a) from EUROSTAT (https://ec.europa.eu/eurostat/data/database):
  - Government revenue, expenditure and main aggregates (gov_10a_main)
  - General government expenditure by function (COFOG) (gov_10a_exp)
  - Main national accounts tax aggregates (gov_10a_taxag)
  - Annual financial accounts for general government (gov_10a_ggfa)
  - Government deficit/surplus, debt and associated data (gov_10dd_edpt1)
  - Transition from the deficit/surplus to the change in debt (gov_10dd_edpt3) GOVERNMENT SUBMODULE 12 / 16
- General government accounts from the Annual National Accounts of the OECD (https://stats.oecd.org/):
  - Taxes and social contributions receipts
  - Government expenditure by function (COFOG)
  - Government deficit/surplus, revenue, expenditure and main aggregates
- Financial Accounts of the OECD (https://stats.oecd.org/):
  - 610. Financial accounts - consolidated - SNA 2008
  - 710. Financial balance sheets - consolidated - SNA 2006
- Supply-and Use Tables from the Annual National Accounts of the OECD (https://stats.oecd.org/):
- 30. Supply at basic prices and its transformation into purchasers' prices
- 40. Use at purchasers' prices
- 43. Use at basic prices
- 44. Valuation matrices
- OECD Inter-Country Input-Output (ICIO) Tables (https://www.oecd.org/sti/ind/inter-country-inputoutput-tables.htm)

- Global Debt Database form the International Monetary Fund https://www.imf.org/external/datamapper/datasets/GDD)

4.4.4.3. DATA FOR THE FIRMS/PRODUCTION SUB-MODULE

The core data of the firms/production submodules is the WILIAM MRIO database. This database includes a set of MRIO tables, socioeconomic and environmental accounts which have been built combining information from the ICIO, EXIOBASE and Eora databases. In particular, the EXIOBASE and Eora datasets have been used to disaggregate some of the industries of the ICIO. Accordingly, we have proceeded to disaggregate in the ICIO those industries more relevant for the analysis of decarbonisation pathways such as primary industries, energy sector or transportation. This facilitates the connection with the land, the materials and the energy modules. The outcome of the whole process is a classification that eventually includes 62 sectors (see Table 1).

The information of the WILIAM-MRIO database has been complemented with data from the EU KLEMS and WORLD KLEMS databases. This information has been used to estimate/calculate some parameters of the firms/production submodule.

Table 1. WILIAM and NACE Rev. 2 sectors.

WILIAM		NACE REV.2
CODE	DESCRIPTION	CODE
1	Crops	011 + 012 + 013
2	Animal production, hunting and related service activities.	014 + 015 + 016 + 017
3	Forestry, logging and related service activities	02
4	Fishing, operating of fish hatcheries and fish farms; service activities incidental to fishing	03
5	Mining of coal and lignite; extraction of peat	05
6	Extraction of crude petroleum and services related to crude oil extraction, excluding surveying	061
7	Extraction of natural gas and services related to natural gas extraction, excluding surveying	062
8	Extraction, liquefaction, and regasification of other petroleum and gaseous materials	091
9	Mining and manufacturing of uranium and thorium ores	0721, part of 24
10	Mining and manufacturing of iron ores	071, part of 24
11	Mining and manufacturing of copper ores and concentrates	Part of 0729, part of 24
12	Mining and manufacturing of nickel ores and concentrates	Part of 0729, part of 24
13	Mining and manufacturing of aluminium ores and concentrates	Part of 0729, part of 24
14	Mining and manufacturing of precious metal ores and concentrates	Part of 0729, part of 24
15	Mining and manufacturing of lead, zinc and tin ores and concentrates	Part of 0729, part of 24
16	Mining and manufacturing of other non-ferrous metal ores and concentrates	Part of 0729, part of 24
17	Quarrying on sand, stone and clay; mining of chemical and fertilizer minerals, production of salt, other mining and quarrying n.e.c.	08
18	Food products, beverages and tobacco	10 + 11 + 12
19	Wood and products of wood and cork; paper products and printing	16 + 17 + 18
20	Manufacture of coke oven products	191
21	Petroleum Refinery	192
22	Chemicals and pharmaceutical products	20 + 21
23	Rubber and plastic products	22
24	Other non-metallic mineral products	23
25	Hydrogen production	Part of 20
26	Fabricated metal products	25
27	Computer, electronic and optical products	26
28	Electrical equipment	27
29	Machinery and equipment, n.e.c.	28
30	Motor vehicles, trailers and semi-trailers and other transport equipment	29 + 30
31	Other manufacturing, including manufacture of furniture, textiles, wearing apparel, leather and related products; repair and installation of machinery and equipment.	13 + 14 + 15 + 31 + 32 + 33
32	Production of electricity by coal	Part of 3511
33	Production of electricity by gas	Part of 3511
34	Production of electricity by nuclear	Part of 3511

35	Production of electricity by hydro	Part of 3511
36	Production of electricity by wind	Part of 3511
37	Production of electricity by petroleum and other oil derivatives	Part of 3511
38	Production of electricity by solar photovoltaic	Part of 3511
39	Production of electricity by solar thermal	Part of 3511
40	Production of electricity n.e.c., including production of electricity by biomass and waste, production of electricity by tide, wave, ocean and production of electricity by Geothermal	Part of 3511
41	Transmission, distribution and trade of electricity	3512 + 3513 + 3514
42	Manufacture of gas; distribution of gaseous fuels through mains	352
43	Steam and hot water supply	353
44	Collection, purification and distribution of water; sewerage; waste collection, treatment and disposal activities; materials recovery; remediation activities and other waste management services	36 + 37 + 38 + 39
45	Construction	41 + 42 + 43
46	Wholesale and retail trade; repair of motor vehicles	45 + 46 + 47
47	Transport via railways	491 + 492
48	Other land transport	493 + 494
49	Transport via pipelines	495
50	Sea and coastal water transport	501 + 502
51	Inland water transport	503 + 504
52	Air transport	51
53	Accommodation and food services	55 + 56
54	Telecommunications, IT and other information services	60 + 61 + 62 + 63
55	Financial and insurance activities	64 + 65 + 66
56	Real estate activities	68
57	Other business sector services, including mining support services and supporting and auxiliary transport activities.	52 + 53 + 69 + 70 + 71+ 72 + 73 + 74 + 75 + 77 + 78 + 79 + 80 + 81 + 82 + 99
58	Public admin. and defence; compulsory social security	84
59	Education	85
60	Human health and social work	86 + 87 + 88
61	Arts, entertainment, recreation and other service activities, including publishing, audiovisual and broadcasting activities	58 +59 +90 + 91 + 92 + 93 + 94 + 95 + 96
62	Private households with employed persons	97 + 98

- n.e.c.: not elsewhere classified
- Source: Own elaboration.

4.4.4.4. DATA FOR THE INTERNATIONAL TRADE SUBMODULE

For the case of parametrization of the international trade submodule the data for estimating Armington elasticities have been collected from the World Input Output Database (WIOD). The WIOD comprises as set of time series (2000-2014) of supply, use and MRIO tables for 56 industries (and 56 products in the case of the supply and use tables) and 43 countries plus the rest of the world as an individual region. The WIOD also contains a set of socioeconomic accounts including, among other information, data on deflators. In the next lines we will detail the data treatment for imports and domestic consumption as well as for deflators.

We use the following information from the cited database:

1) International use tables from the WIOD: these tables provide, for each of the users in the Input-Output (IO) framework (industries and final demand components), information on the intermediate and final use of domestic and imported products expressed in millions of current US dollars. In the case of imported products, the use tables report the country of origin and in "free on board prices" (fob) prices. These prices have been transformed to "cost insurance and freight" (cif) prices by distributing the row of "International trade and transportation margins" proportionally across imported goods and countries.

2) Deflators Socio-Economic Accounts from the WIOD: we have also used the deflators of the Gross Domestic Output (GO_PI) and the Value Added (VA_PI) for the creation of the price variables. The

deflators of the Gross Domestic Output have also been used for deflating the intermediate and final use of domestic and imported products from the WIOD and ICIO datasets. These deflators are reported by industry and have been transformed to product deflators, using de markets shares of the supply table, in order to deflate the use tables of WIOD which are product by industry. Note that these deflators are only available for the 43 WIOD countries, therefore the analysis with the ICIO database will be limited to the 43 countries that are included in both databases.

Information on sectoral production Socio-Economic Accounts from the WIOD: Data on intermediate inputs and gross output production by sector, country and year has been obtained from Socio Economic Accounts. As this data is expressed in local currency in local terms, it has been converted to constant U.S dollars for the year 2000.

For bilateral trade data (gravity equation) we have used data from WIOD as well as from ICIO(OECD) and deflation procedure as for the estimation of Armington elasticities, although now the prices of imports are country specific (for the first node we used a single price per product). For prices, we will employ the information already created in the first node for the sectoral domestic price as well as for the aggregated import price.

Regarding the data sources and data collection for the additional covariates for the gravity equation data regarding GDP, GDP per capita, Free trade agreements, common official language, tariffs and distance have been obtained from a squared gravity dataset available at the Centre d´Études Prospectives et d´Information Internationales (CEPII) (Head & Mayer, 2013).

4.4.4.5. DATA FOR THE LABOUR SUBMODULE

Regarding the labour market submodule, the three main data sources for the econometric analysis are: (i) EUKLEMS database (2019 release), (ii) EUROSTAT and (iii) OECD. EUKLEMS database (Stehrer et al., 2019) provide industry level data for hourly productivity of employees, hourly compensation to employees and socio demographic variables. offering data for the EU27 countries, Japan and USA. National accounts compile data on industry level real gross output (GO_Q in EUKLEMS code), number of hours worked by employees (H_EMPE) as well as compensation of employees (COMP) at current prices among other variables of interest from the year 1995 up to 2017. Labour accounts of EUKLEMS offer data on employment share over the active population by socio demographic groups, namely, young, medium age, old, low-skilled, medium-skilled, high-skilled, female and male workers. In this case, data goes from 2008 up to 2017 although Belgium and Japan provide data since 1999 and 1995 respectively. Data for the United States is provided at the aggregate level, with data available at the industry level. Regarding data coming from national accounts, industry disaggregation in higher than it is for labour accounts. On the other hand, there is no data availability on real gross output for UK, Bulgaria, Cyprus, Spain, Ireland, Malta and Lithuania. Furthermore, variables coming from Labour accounts only present data for aggregated industries while national accounts provide a wider industry disaggregation for manufacturing and service-related activities. National accounts provide data on 43 industries while labour accounts compiles to 18 industry categories in an upper level of aggregation.

For the unemployment term, data for effective unemployment rate and natural rate of unemployment (NAIRU) are obtained from the AMECO database at EUROSTAT, that covers relevant databases for macroeconomic modelling of the EU countries. In this case, annual data for EU27 countries and non-EU countries as Montenegro, North-Macedonia, Serbia, Turkey, Iceland, Norway, Switzerland, Australia, New Zealand, Australia and United States is provided for the years 1995 to 2019. Besides that, other general aggregate variables are used. That comprises the inflation rate or consumer price index (for $P_{k,t}^{CP}$) which is obtained from OECD database (Consumer price indices (CPIs) - Complete database, (https://stats.oecd.org) and goes from the year 1995 to 2018. This last variable is specified as an index with base or reference year being set up at 2015.

4.4.5. IMPLEMENTED POLICIES

In this section we list some of the policies parameters of the economic module.

- $Exo\_Debt\_Interest\_rate_{rt}$: interest rate paid for debt (units: rate).
- $Exo\_Gov\_DeficitSurplus\_to\_GDP\_obj_{rt}$: exogenous Government deficit/surplus to the GDP.
- $Exo\_Hours\_worker_{rit}$: average hours worked per worker (units: millions).
- $Exo\_Tax\_rate\_Operating\_Surplus_{rit}$: tax rate on operating surplus (benefits) (units: rate).
- $Exo\_Tax\_rate\_output_{rit}$: net tax rate on output (units: rate).
- $Exo\_Tax\_rate\_products\_domestic_{ric0}$: tax rate on products of the base year (t=0) (units: rate).
- $Exo\_Tax\_rate\_products\_domestic_{riut}$: tax rate for domestic products (units: rate).
- $Exo\_Tax\_rate\_products\_import_{sriut}$: tax rate for imported products (units: rate).
- $Exo\_Tax\_rate\_Income_{rht}$: tax rate on income by region, household type and year (units: rate).
- $Exo\_Tax\_rate\_Wealth_{rht}$.: tax rates on wealth respectively (units: rate).

4.4.6. KEY OUTPUTS

This Sections present a summary of the key outputs of the model. Most variables in the economic module are expressed in million USD or in million USD at 2015 constant prices (those expressed in variables in real terms, with the suffix $\_real$).

- $Change\_Gov\_Debt_{rt}$ is to the change in the stock of gross debt.
- $Capital\_Damage\_CC_{rit}$: effect of capital consumption due to damages.
- $Capital\_stock\_real_{rit}$: actual capital stock in real terms.
- $Consumption\_Appliances_{rht}$: consumption of appliances.
- $Consumption\_Durable_{rht}$: consumption of durable goods by region.
- $Consumption\_HH\_Energy_{rhft}$: household energy demand by fuel.
- $Consumption\_Housing\_services_{rht}$ consumption of housing services.
- $Consumption\_Non\_Durable\_Energy_{hrt}$: expenditure on energy and transportation.
- $Consumption\_Non\_Durable\_Non\_Energy_{rht}$: total household consumption of non-durable non-energy goods.
- $Consumption\_Non\_Durable_{rht}$: consumption of non-durable goods.
- $Consumption\_Private\_Transport_{rhft}$: expenditure in current prices for private transport.
- $Consumption\_Private\_transport_{rht}$: total expenditure for private transport.
- $Consumption\_Public\_transport_{rhmt}$: expenditure in current prices for public transport.
- $Consumption\_Public\_transport_{rht}$: total expenditure in public transport.
- $Consumption\_Total_{rht}$ total consumption.
- $Consumption\_Vehicles\_Total\_real_{rt}$: total consumption of vehicles in real terms.
- $Consumption\_Vehicles_{rht}$: consumption of vehicles.
- $Consumption\_by\_category_{rhlt}$: consumption by category.
- $Disposable\_income_{rht}$: disposable income.
- $Employment\_total_{rt}$: total employment (units: thousands).
- $Employment_{rit}$: employment (units: thousands).
- $Exports\_real_{rsit}$: exports in real terms.
- $Final\_demand\_BP\_real_{rict}$: final demand in real terms at basic prices.
- $Final\_demand\_PP\_real_{rict}$: final demand in real terms at purchaser prices.
- GDP_real_{rt}: Gross Domestic Product in real terms.
- $GFCF\_PP\_real_{rit}$: real gross fixed capital formation at purchaser prices.
- $Gov\_Deficit\_Surplus\_to\_GDP_{rt}$: ratio of government deficit/surplus to GDP (units: rate).
- $Gov\_Deficit\_Surplus_{rt}$: government actual deficit/surplus at the end of the period.

- $Gov\_Expenditure_{rt}$: actual value of government expenditure at the end of the period.
- $Gov\_revenue_{rt}$: government revenue.
- $Gross\_Operating\_surpls\_real_{rit}$: gross operating surplus in real terms.
- $Gross\_value\_added\_real_{rit}$: gross value added in real terms.
- $Imports\_real_{rsit}$: imxports in real terms.
- $Labour\ compensation\_real_{rit}$: labour compensation in real terms.
- NDP_real_{rit}: Net Domestic Product.
- $Output\_real_{rit}$: output in real terms.
- PCP_{rt}: consumer price index (units: index).
- $Price_{rit}$: production price (units: index).
- $Unemployment\_rate_{rt}$: unemployment rate (units: rate).

4.5. THE ENERGY MODULE

4.5.1. GENERAL DESCRIPTION

The WILIAM Energy module encapsulates the entire transformation process from primary to final energy necessary to fulfil societal economic demands. This module is parameterized across nine geographical regions and integrates seven final energy commodities, 12 primary energy commodities and 35 energy transformation technologies for electricity and heat.

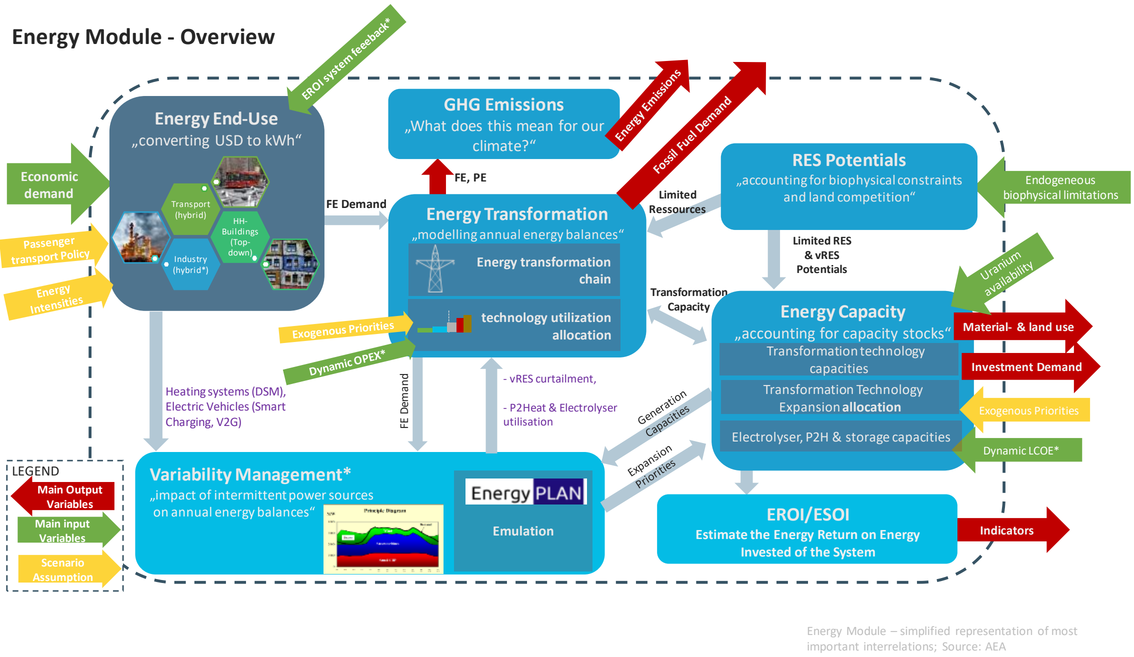


Figure 16: Energy Module – Overview.

The model comprises seven closely interrelated submodules (see Figure 16 above and detailed descriptions in section 4.5.2):

1. The End-Use submodule translates the economic demand (in USD) from the Economy module into final energy demand (in energy units, distinguished between different final energy commodities). It uses comprehensive bottom-up models to represent passenger transport energy demand and the energy transformation chain. Conversely, it computes energy demand from the industry in a top-down approach using energy intensity data.

2. The Energy Transformation submodule processes the final energy demand from the End-Use submodule and calculates the resulting primary energy demand. It considers transmission and storage losses, sector energy self-consumption, transformation losses, and refinery losses. This submodule includes a detailed allocation of the required transformation output to the 35 modelled transformation technologies for heat and electricity, significantly influencing the short-term dynamics of the energy system.
3. The Energy Capacity submodule keeps track of the installed transformation, storage, Power2Heat, and electrolyser capacities. The allocation of new capacity to the different technologies is a major factor influencing the long-term dynamics of the energy system.
4. Capacity expansion is bound by constraints such as techno-economic potentials, uranium availability or land-use availability modelled in the RES Potentials submodule.
5. The Variability Management submodule approximates the effects of increasing shares of intermittent renewable energy generation on the power system by emulating results from the dispatch model EnergyPLAN (Lund et al., 2021).
6. The EROI / ESOI submodule is computed from the bottom-up material requirements (cf. section 4.5.2.4) and is designated for the dynamic calculation of energy return on investment at technology and system levels.
7. The GHG Emissions submodule is included for estimating greenhouse gas emissions resulting from energy production.

The outcomes of the energy module are primarily driven by the economic demand simulated by the Economy module (see Section 4.4) considering restrictions from other parts of WILIAM (e.g., land-use).

4.5.2. SUBMODULES

4.5.2.1. END-USE SUBMODULE

The End-Use submodule translates the economic demand generated by the Economy module into the final energy demand[13]. Two main approaches are commonly used in the literature for the modelling of future energy demand: (1) through the modelling of bottom-up, detailed demand for end-use services (e.g., passenger per km, freight per km; m^2 of buildings, kg of steel or fertilizer etc.); and (2) top-down through the projection of general trends at sectoral level applying econometrics to time series of aggregated indicators such as GDPpc, population, energy intensities, physical scarcities or prices.

In the WILIAM model, the future energy demand is modelled combining these approaches: (i) a top-down approach through final energy intensities for non-energy sectors and for the non-energy uses and (ii) a bottom-up approach to estimate the final energy demand of energy sectors (ii.a) and households' transport (ii.b) and industrial hydrogen (ii.c) [14].

(i) The energy demand for non-energy sectors is calculated using a top-down approach through final energy intensities. To obtain energy intensities, data from OECD (IEA, 2019) and EXIOBASE (Stadler et al., 2019) databases with their satellite accounts have been used. The energy carriers available in these databases have been aggregated into 7 types of final energy: electricity, heat, hydrogen, liquids, gases, solids fossil and solids bio. Energy intensities are obtained individually for each of the non-energy sectors. The dynamic of each of the sectoral final energy intensity is broken down into (1) improvement in energy efficiency and (2) final energy substitution. The speed of changes in these factors depends mainly on historical trends and supply-demand unbalances in the market (de Blas et al., 2019). The energy demand

[13] Including non-energetic use of fossil fuels

[14] The bottom-up representation will be extended by and (ii.c) households' buildings and (iii.d) key energy intensive industrial sectors like iron and steel and chemistry in future versions.

for non-energy sectors is finally obtained by multiplying the economic output per sector obtained in the economy module and the final energy intensity per sector and final energy.

(ii.a) The final energy demand of the energy sectors is modelled using a bottom-up approach in the energy transformation submodule, which calculates the energy self-consumption of the energy sectors, the transformation inputs, the transformation outputs (being the difference between the latter two the transformation losses) and the transport and distribution losses (see section4.5.2.2).

(ii.b) The final energy demand of households' transport is modelled using a bottom-up approach through physical demand (total demand for transportation in million passenger-kilometres) obtained in the economy module. The bottom-up transport submodule computes, among other variables, the demand of public transportation by mode and the physical energy use for private transportation by fuel. WILIAM differentiates between 9 types of passenger transport modes: LDV (light duty vehicle), bus, motorbikes, rail, air_domestic, air_intraEU, air_international, marine and NMT (non-motorised transport) and 10 power trains: ICE gasoline (gasoline internal combustion engine), ICE diesel, ICE gas, ICE LPG (liquefied petroleum gas), BEV (battery electric vehicle), PHEV (Plug-in hybrid electric vehicles), HEV (hybrid electric vehicles), FCEV (Fuel cell electric vehicles), EV (other electric vehicles) and HPV (Human-powered vehicles). The behavioural changes variables modify the total transport demand coming from the economic module through modal share and vehicle fleet demand with load factor variable as can be seen in Figure 17.

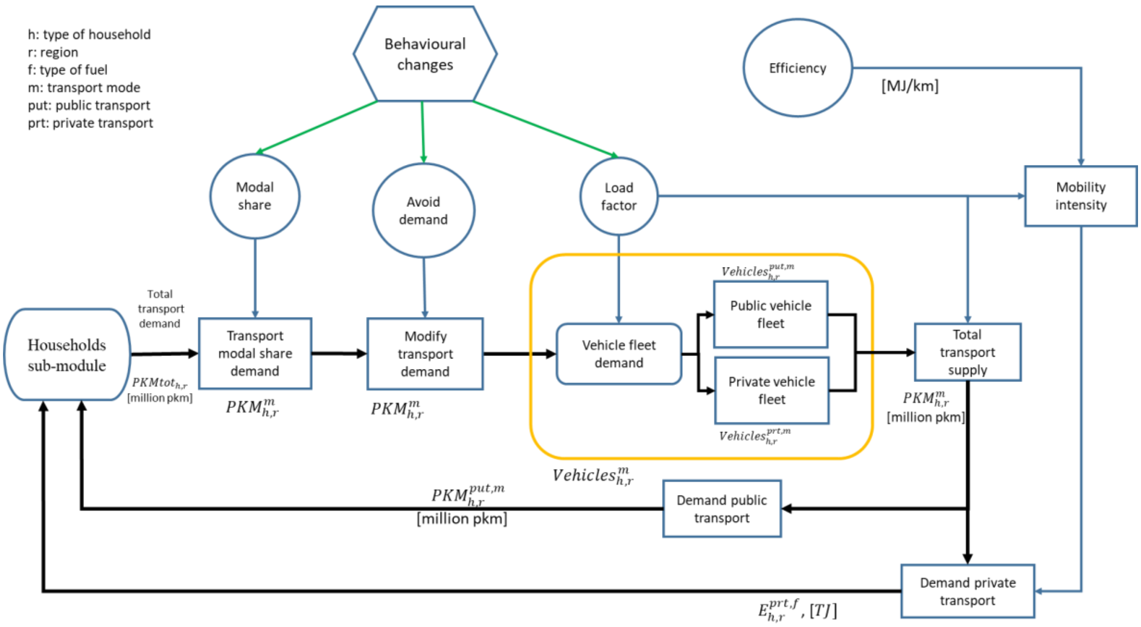


Figure 17: Internal connections of the households' transport submodule.

(ii.c) The final energy demand of the hydrogen industrial sector is modelled using a bottom-up approach from the energy/material/economic requirements from different sub-modules. Industrial uses of hydrogen include refining liquids (oil and biodiesel), ammonia and methanol production and steel production. Intensities have been calibrated to determine the amount of hydrogen required for each of the hydrogen demanding sectors. This calibration has been done on the basis of historical data or typical operating parameters, physical limits of the process, etc. justified in the literature.

4.5.2.2. TRANSFORMATION SUBMODULE

The energy module, at its heart, consists of the transformation submodule, which is designed to generate consistent simulations of annual energy balances for the nine major regions. Figure 18 gives an overview of the most important calculation steps in the transformation module. It receives final energy demand data from the End-Use submodule ("final_energy_demand_by_FE_9R") and estimates the primary energy demand required to meet this final demand ("PE_by_commodity"). It does so by considering multiple factors including transmission- and storage losses ("PROSUP_transmission_losses", "PROSUP_storage_losses"), sector energy's own consumption ("PROSUP_sector_energy_own_ consumption_per_commodity"), the additional electricity required for flexibility processes P2H and electrolysers losses in transformation for heat and electricity ("PROTRA_conversion_efficiencies"), and refineration losses ("PROREF_CONVERSION_FACTORS").

While incorporating energy losses related to refineration processes, it does not explicitly account for refineration capacities, hence implicitly assuming that there will be adequate refineration capacities at any given time. This strong simplification of the current energy system seems justified given that WILIAM was designed to explore future carbon neutral scenarios where these technologies will only play a minor role.

The submodule further receives an estimation of potential RES curtailment brought about by high penetrations of photovoltaics (PV) and wind energy in the system by the variability management submodule ("protra_max_full_load_hours_curtailed", see subsection below).

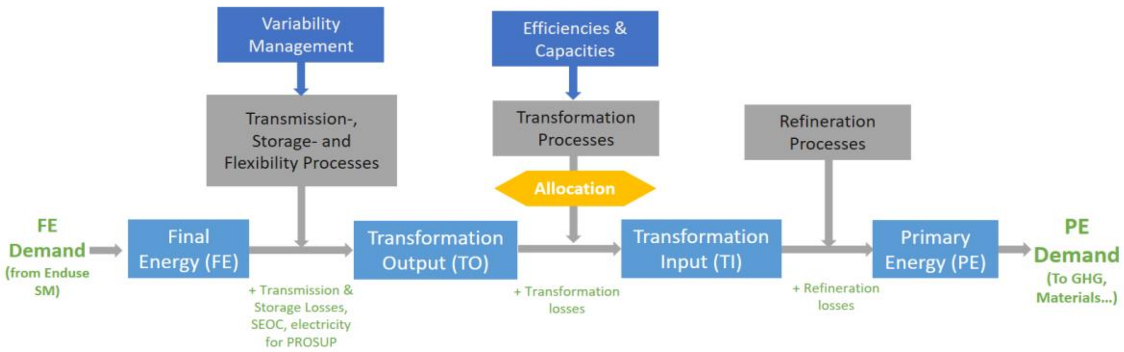


Figure 18: Flow of calculations in the transformation submodule. A complete figure with all processes is given in Figure 19 and Figure 20.

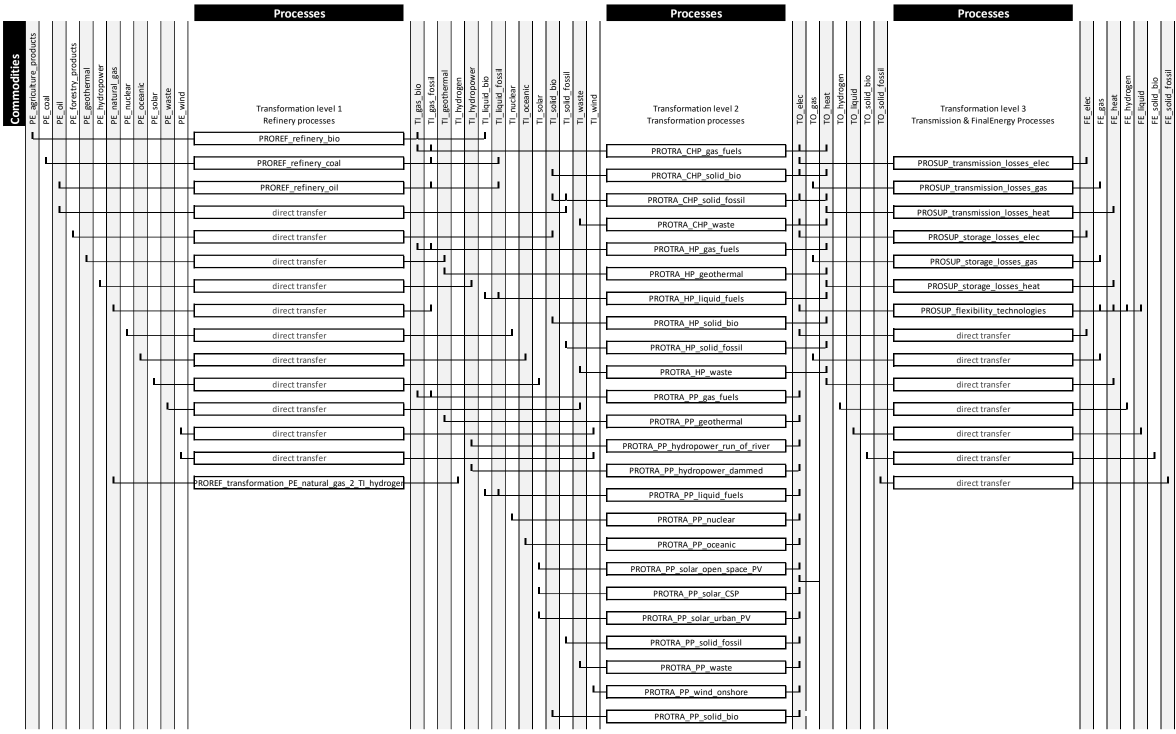


Figure 19: Flow detailed process diagram of all processes and flows in the transformation submodule part 1.

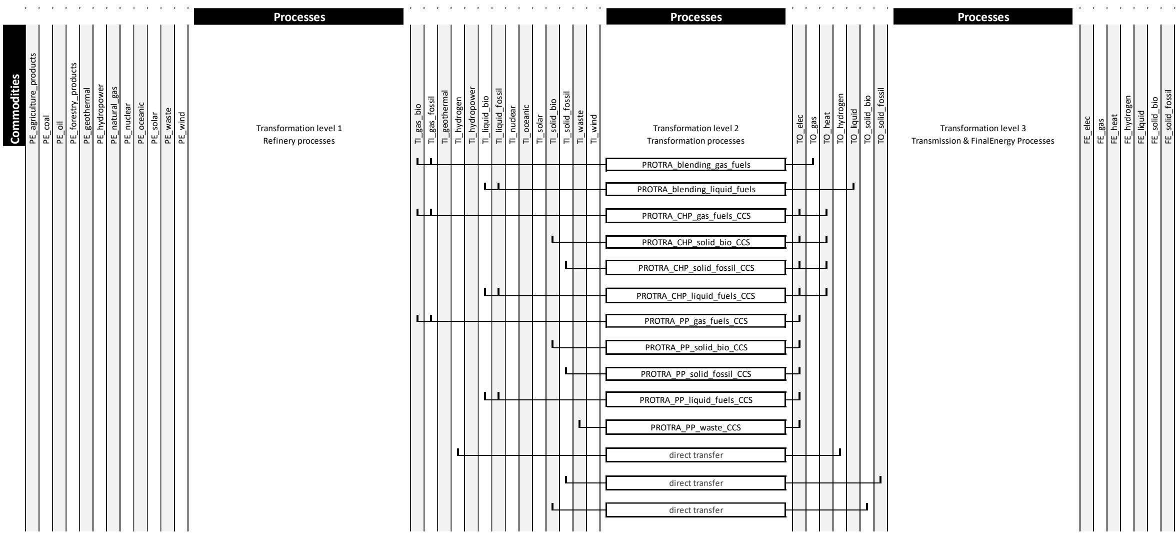


Figure 20: Flow detailed process diagram of all processes and flows in the transformation submodule part 2.

One essential part of the submodule resides in the transformation process allocation, where the necessary quantity of electricity and heat (TO_elec and TO_heat) is apportioned to the 35 distinct categories of transformation processes (PROTRA). WILIAM in general distinguishes three types of transformation technologies:

- Power Plants (PP) that generate electricity,
- Heatplants (HP) that exclusively produce (district-) heat, and
- Combined heat and power plants (CHP) that simultaneously produce both heat and power.

This significantly affects the short-term dynamics of the model. While in optimization models like e.g., ETSAP-TIMES allocation is the result of a cost-based solving algorithm (Loulou et al., 2005), simulation models like WILIAM require to explicitly model the causal relationship and implicit logic of the underlying processes.

In light of this, the allocation process was modelled in a bifurcated process (Figure 21): First the heat demand is distributed among HP and CHP technologies. In a second step the electricity produced by CHPs (assuming a constant heat-to-power ratio for CHPs, derived from empirical IEA data) is subtracted from the aggregate electricity demand. Finally, the residual demand is allocated to PP technologies.

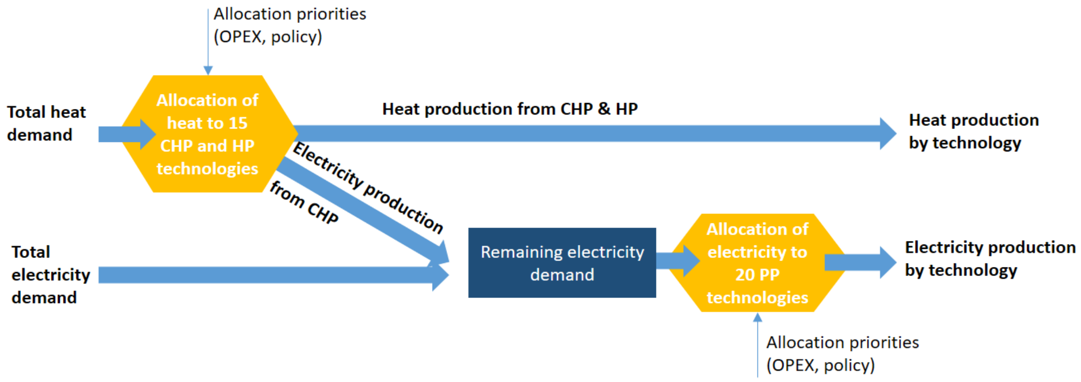


Figure 21: Stepwise allocation of total (district-) heat and electricity demand to the CHP, Heat Plant (HP) and power plant (PP) technologies.

This stepwise allocation approach utilising constant heat-to-power ratios provides only a rudimentary and simplified representation of CHP operating strategies. This method indirectly assumes a heat-demand driven operation strategy for all CHPs. Although it is reasonable to assert that CHPs are designed to optimise fuel efficiency in temperate climates (e.g., in regions like India or Latin America district heat plays only an insignificant role), the actual CHP operational strategies implemented by energy utilities are considerably more intricate. These strategies may fluctuate annually and are influenced by a myriad of factors, such as price ratios between electricity, heat, and input fuel, grid congestions or the prevailing subsidies and regulations within the country. Consequently, this structure might contribute to some unavoidable discrepancies between modelled outcomes and empirical data. Moreover, unreliable data quality, particularly within the heat sector, adds another layer of complexity[15]. These uncertainties and methodological aspects need to be kept in mind when interpreting the results of the module.

For the allocation, WILIAM uses a “one-to-many” allocation method, allocating one quantity (e.g., required electricity output) to a vector of different options (e.g., maximum available PP output) according to given priorities. While the algorithm is complex and offers a set of parameters[16] to shape the degree of response to changing priorities, the general idea is that technologies with higher priority are utilized to a higher degree than technologies with lower priority.

The allocation priority of the PROTRA technology utilization allocation can be set to (a) endogenous, (b) exogenous/policy or (c) a mix of the before mentioned. The endogenous signal is based on a calculation of the operation expenditures like displayed in the following equation.

[15] While data related to the power sector are generally well-documented and largely reliable, the heat sector receives significantly less attention. This discrepancy holds true for both transformation output and, more critically, for installed capacities.

[16] The Vensim® function “ALLOCATE_AVAILABLE” uses “pwidth” and “ptype” parameters, see https://www.vensim.com/documentation/fn_allocate_by_priority.html for details.

$$OPEX_{R9,PROTRA} = \frac{1}{\eta_{R9,PROTRA}} * (p\_fuel_{PROTRA} + EF_{PROTRA} * p\_CO2_R) + O\&M_{PROTRA}$$

Equation 109

OPEX (by region and technology) resembles a simplified estimation of operation expenditures in USD/MWh, where:

- η is the dimensionless conversion efficiency of each transformation technology by region,
- p_fuel is the fuel cost of the technology, e.g., gas cost for gas power plants, endogenously coming from the materials module (for all fossil fuels) in USD/MWh,
- EF is the emission factor of the respective technology in tCO2/MWh,
- p_CO2 is the CO2 price per region (exogenous) in USD/tCO2, and
- O&M is the operation and maintenance expenditure by technology in USD/MWh.

For renewable technologies like PV and wind, fuel cost and co2 cost will be zero, and hence the OPEX will consist only of the O&M element. The most essential dynamic in this equation comes from the fossil fuel cost, which is calculated endogenously in the Materials module on a global basis (see section 4.9). Note that only if set to endogenous the feedback loop with the materials- and economy module is fully closed.

Note that transformation technologies based on biomass (CHP, PP, and HP_solid_bio) we do not have an endogenous unit price in the model. However, we do have an availability factor form the Land and Water module that can serve as a scarcity feedback variable (see section 4.7 for details). The approach taken for slolid_bio was to start off from an supply cost curves (Rogner, Aguilera, Archer, Bertani, Bhattacharya, Dusseault, Gagnon, Haberl, Hoogwijk, Johnson, Rogner, Wagner, Yakushev, et al., 2012) and increase the price in case of scarcity by dividing it by the availability factor (number between 1 and 0, 1 meaning that 100% percent of the demand can be met).

Also, for nuclear the model has no endogenous price mechanism available: However, the operation logic of nuclear is different from other energy carriers, as the employment of this technology seems to be a far more political issue and less subject to cost-benefit analyses. Hence the simplified assumption was taken that, if a region possessed nuclear power capacities, they will always be utilized to their full extent (by giving it the highest priority in the allocation). However, to avoid unrealistic scenarios the build-up of nuclear capacities is restricted by the capacity submodule (see section 4.5.2.3 below).

With regard to hydrogen, the installed capacity of electrolysers (stationary + flexible), depending on their capacity factor, results in the generation of a certain amount of hydrogen and electricity consumption (in each of the nine regions), both of which are placed in the energy chain in the level of Transformation Output (TO). As for the hydrogen generated by the electrolysers it can be used in its pure form ('TO_pure_hydrogen'), as hydrogen liquids-based fuel ('To_H2_liquids_based_fuel') or as hydrogen gases-based fuel ('TO_H2_gases_based_fuel'), the latter two accounting for the energy loss associated with the inefficiency of the process (H2 -> H2 liquid fuel and H2 -> H2 gas fuel). The proportion of the hydrogen generated that is dedicated to the production of synthetic fuels (liquids-based fuel + hydrogen gases-based) is given in the form of liquids and gases substitution policies[17] that are used as final energy. The Total electricity consumed to generate hydrogen (pure hydrogen + hydrogen used to produce synthetic fuels) is collected in the variable 'TO_elec_consumption_stationary_electrolyzers_by_PROSUP_H2'. As a result, the commodities associated with the hydrogen sector are collected in the variable 'TO_commodities_H2', which includes energy in the form of electricity consumed by the electrolysers

[17] The policies of substituting hydrogen-based fuels for liquids and gases are part of the 8 alternative policies used as inputs to the WILIAM model explorer.

(subscript 'TO_elec') and energy in the form of pure hydrogen (subscript 'TO_hydrogen'), hydrogen liquids-based fuel (subscript 'TO_liquid') and hydrogen gases-based fuel (subscript 'TO_gas'). Electricity consumption has a positive sign (to increment the commodity in the transformation chain) and the rest has a negative sign (to subtract what is already produced). Finally, the variable 'TO_commodoties_H2' is included in the transformation chain by the variable 'PROSUP_flexibility_technologies', which collects the energy in the form of Transformation Output of the different flexibility technologies. This last variable is added to the variable 'TO_by_commodity', thus determining the total energy in the form of Transformation Output.

4.5.2.3. CAPACITY SUBMODULE

The main function of the capacity module is to keep track of the PROTRA and PROFLEX capacity stock (for heat and electricity) in accounting for decommissioned capacities and newly added capacities. Owing to the capacity module's role in dictating the availability of transformation technologies, this submodule exerts a significant influence on the long-term dynamics of the power system.

To enhance clarity without adding undue complexity, the capacity submodule adheres to a simplified method. Herein, a fraction equivalent to 1/lifetime of each existing capacity is phased out annually. The emerging shortfall of transformation capacity output - including any deficit potentially originating from energy demand growth- is allocated among the different PROTRA technologies, taking into consideration both capacity expansion priorities and biophysical limitations. At the core of the submodule is the distribution of the newly required transformation output (heat and electricity) to the 35 PROTRA technologies. Drawing parallels with the allocation in the transformation submodule, this allocation proceeds in a stepwise manner, first allocating heat to HP and CHP technologies, and subsequently allocating the remaining electricity to PP technologies.

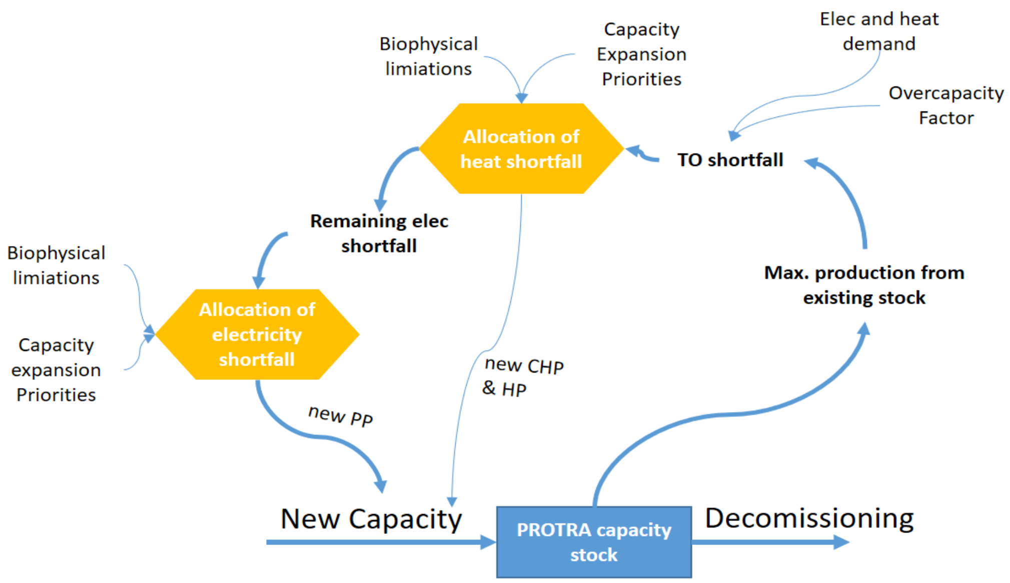


Figure 22: Simplified structure of the capacity submodule and the allocation mechanism.

The capacity stock is fixed to empirical values until 2020, following which expansion dynamics are activated. The capacity data was calibrated in line with various international data sources and was supplemented by regional sources to rectify any implausible data points.

Regarding the allocation priorities a mixed endogenous / exogenous approach was chosen (similar to the technology utilization allocation in section 4.5.2.2). The endogenous part of the allocation priority signal is based on a simplified calculation of life cycle cost of electricity (LCOE). This simplified representation of LCOE that ignores interest rates, discount rates, and eventual reinvestments during the lifetime of the technology.

It is calculated according to the following formula:

$$LCOE_{R9,PROTRA} = \frac{Lifetime\ Cost}{Lifetime\ Output} = \frac{I_{R9,PROTRA} + OPEX_{R9,PROTRA} * LT_{R9,PROTRA} * FLH_{R,PROTRA}}{FLH_{R,PROTRA} * LT_{R9,PROTRA}}$$

Equation 110

- I is investment in USD/MW capacity installed
- LT is the lifetime of the technology in years
- OPEX is the (simplified) operating expenditure in USD/MWh (See Equation 109)
- FLH is the full load hours in hours per year (Equivalent to capacity factor * 8760 h/year)

Two important endogenous feedbacks affect this variable: 1) the endogenous OPEX calculation depending on fuel price and CO2 price and 2) the full load hours of intermittent renewable technologies (PV and Wind) that might reduce due to curtailment in high RES scenarios (see section 4.5.2.6).

However, it can be argued that the choice of new energy transformation capacities is not only driven by economic reasoning, but also subject to political factors (e.g., no new nuclear reactors would be built if cost-benefit analysis would be the only viable criteria), which is why the user can also define capacity expansion priorities exogenously and assign them a certain policy weight.

For the BAU scenario, the exogenous expansion priority parameter was estimated through an iterative process: (1) calculation of empirical average PROTRA-shares by region for the period 2017-2019, (2) normalisation of the shares within the 0 (smallest share) to 1 (largest share) range, (3) manual increase of vRES priorities (solar and wind) to compensate for their underrepresentation in current capacity and the strong growth trends of these technologies in current trends, (4) model execution and results observation, and (5) a return to step 3 for further adjustments, if necessary.

The scope for expanding renewable capacities is limited by the RES potentials submodule (see section 4.5.2.5. The build-up of new fossil capacities remains unconstrained in this version; and for nuclear power plants there is a dependence on uranium availability.

This module allows computing the total capacity investment cost ($) of PROTRAS + PROFLEX over time for the 9 regions. These are obtained after having collated data of unitary capacity investment costs ($/MW) for these technologies and considering uncertainties in the projections. The data have been converted to 2015 US$ for consistency and in the model these prices are endogenously updated by the price of investments from the economy module ("price_GFCF"), which also allows to consistency compare these investments with the level of GDP for each region. These calculations are an intermediary step towards endogenizing the investments (GFCF) of energy sectors in the economy module.

To improve the usability of the module and given the existing uncertainties with relation to non-commercial energy technologies, a SWITCH to differentiate between currently mature and commercial technologies and those which today are non-commercial has been introduced ("SELECT_AVAILABILITY_UNMATURE_ENERGY_TECHNOLOGIES_SP"). The today non-commercial technologies are electrolysers, CCS plants, wind offshore floating and marine. This switch allows the user to run simulations allowing these technologies to be present in the future mix in a customized way.

With regard to hydrogen, the installed capacity of two types of electrolysers has been modelled for 9 regions: stationary ('stationary_electrolyzer_capacity_stock' in TW) and flexible electrolysers ('flexible_electrolysers_capacity_stock' in TW). The former are those used to satisfy the final energy supply, have a predefined capacity factor of 50% and it is assumed that all their hydrogen is produced to be sold on the market (not in-situ production). The latter are those used for flexibility purposes, and their utilization depends on energy variability management. The capacity expansion of the stationary electrolysers is, assuming the SWITCH 'SELECT_AVAILABILITY_UNMATURE_ENERGY_TECHNOLOGIES_SP' is active, endogenous and based on hydrogen demand (pure hydrogen + hydrogen used to produce synthetic fuels); the total hydrogen demand minus the hydrogen generated by the flexible electrolysers is equal to the hydrogen that should be produced by the stationary electrolysers. When the stationary electrolysers are not capable of producing such a quantity of hydrogen, more capacity will be installed. In addition, the capacity stock of stationary electrolysers may be reduced due to the end-of-life of the electrolyser units ('LIFETIME_ELECTROLYZERS'), at the rate between the following variables: 'stationary_electrolyzer_capacity_stock'/'LIFETIME_ELECTROLYZERS'. Regarding the installed capacity of flexible electrolysers, in addition to their decommission taking into account their lifetime (as in the case of stationary electrolysers), the capacity stock can be increased exogenously (SWITCH 'SWITCH_NRG_PROFLEX_CAPACITY_EXPANSION_ENDOGENOUS' ON) based on policies with a certain installed capacity target or endogenously driven by the allocation of curtailment (SWITCH 'SWITCH_NRG_PROFLEX_CAPACITY_EXPANSION_ENDOGENOUS' ON).

4.5.2.4. EROI/ESOI SUBMODULE

The investment profile of RES is very different from their fossil fuel counterparts, requiring substantial up-front costs before delivering gross power. The energy investments for RES and electric batteries to produce electricity are endogenously and dynamically modelled in WILIAM, which allows the Energy Return on Energy Investment (EROI, ESOI for storage) of individual technologies and the EROI of the whole energy system to be computed. The conceptual approach and addressed technologies is the same than the one applied in MEDEAS (Capellán-Pérez et al., 2019, 2020): solar PV, solar CSP, wind onshore, wind offshore and electric batteries, although some updates have been included notably disaggregating by sub-technologies EV batteries (LMO, NCA, NMC622, NMC811, LFP) and solar PV (ground and rooftop; poly-Si, mono-Si, CIGS, CdTe) (Pulido Sánchez et al., 2020, 2021). In terms of boundaries, since grids are not explicitly modelled in WILIAM, we are limited to the standard EROI (EROIst), which includes the on-site and offsite (i.e., energy needed to manufacture the devices and systems used later onsite) energy requirements to get the energy (e.g., build, operate and maintain a power plant).

The EROI for solar PV, solar CSP, wind onshore, wind offshore and electric batteries (cf. Equation 111) is dynamically and endogenously computed combining the material intensities of each technology with the embodied energy of materials -including accounting for the material recycling rates- and technical performance factors (efficiencies, capacity factors, self-consumptions, etc.). Other technologies such as hydropower, geothermal, marine or Pumped Hydro Storage (PHS) are covered in a stylized way through an exogenous EROI ratio (i.e., do not distinguishing between the upfront investment phase and the energy returning one). The EROI of fossil fuels and the energy embodied per kg of extracted materials are assumed constant for the sake of simplicity. The EROI of the system includes additional investments for the functioning of the system which cannot be allocated to one technology, notable those related to variability management such as storage. The EROI of the system (Equation 112) was used to feedback into the energy demand, a feature which is not yet implemented in WILIAM. Different options exist, such as linking directly to the Investment submodule in the Economy module. More details can be found in (Ólafsdóttir et al., n.d.).

$$EROI_i = \frac{Annual\ elec\ output_i \cdot (1 - GCF) \cdot L_i}{EnU_i^{New\ cap+OG} + EnU_i^{Decom\ wear\ cap} + EnU_i^{O\&M} \cdot L_i + Annual\ elec\ output_i \cdot L_i \cdot SC_i}$$

Equation 111

- Where $Annual\ elec\ output_i = CF_i \cdot Installed\ new\ capacity_i \cdot 8760\ \frac{h}{yr}$
- i: electricity generation technology.
- Annual elec output: Annual electricity output.
- CF: capacity factor.
- L: lifetime of the installed infrastructure.
- $EnU^{New\ cap+OG}$: final energy used in the construction of new capacity and overgrids for RES variables.
- $EnU^{Decom\ wear\ cap}$: final energy used for decommissioning those infrastructures that have ended their lifetime. We assume a fixed share in relation to the EnU of the energy required for the construction of each power plant of 10% following (Hertwich et al., 2015), i.e., Decomm=0.1.
- GCF: Annual losses due to the Joule heating within each power plant (grid-correction factor) as a share of total annual electricity output. This is endogenously calculated by the model as distribution and transmission losses.
- $EnU^{O\&M}$: annual energy used for the operation and maintenance.
- SC: electricity self-consumption of the power plant as a share of the electricity output.

$$EROI(t)_{system}^{st} = \frac{TFEC(t)}{OEU(t) + TFEI(t)_{RES\ elec} + TFEI(t)_{storage\ elec}}$$

Equation 112

- TFEC: total final energy consumption (excluding energy materials for non-energy uses).
- $TFEI_{RES\ elec}$: total final energy investments for renewable technologies of electricity generation (construction, replacements, operation and maintenance, decommissioning and overgrids).
- $TFEI_{storage\ elec}$: total final energy investments for storage of electricity
- OEU: Energy industry own energy use

4.5.2.5. RES POTENTIALS SUBMODULE

The modelling RES potentials aims at representing in the model their techno-sustainable potentials, which refers to the part of the technical energy potential that is subject to sustainability criteria, that is the potential that takes into account sustainability, ecological and socio-economic aspects (see Figure 23).

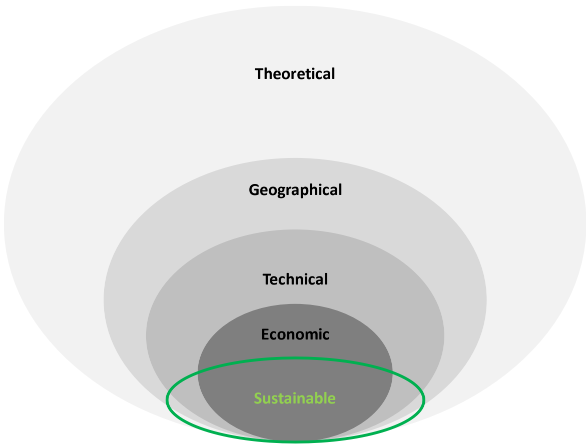


Figure 23: Types of renewable energy potentials. The arrows between the economic and techno-sustainable potential demonstrate that the boundaries between them might be overlapping. Source: own visualisation based on LOCOMOTION D7.2 (Papagianni et al., 2020).

In WILIAM two main cases for the modelling of RES potentials are considered:

1. Exogenously initialized RES potentials in the energy module and modelled independently from the land-use module: estimated out of the model and initialized exogenously in energy units in the energy module (and hence subject to exogenous parameters options), but which can be modified endogenously during the simulation getting information from other parts of the model (technological change, etc.)., Here we can distinguish 2 subtypes:
   i. Those coming from GIS analyses (wind onshore & offshore): different potentials depending on the EROImin criteria applying Dupont's method (Dupont et al., 2018). This method uses a grid-cell approach to assess the theoretical wind potential in all geographic locations by considering technological and land-use (suitability) constraints, as well as a top-down limitation on kinetic energy available in the atmospheric boundary layer is imposed.[18] An analysis is then performed where the EROI of the wind potential is evaluated. With these constraints wind farm designs are optimized in order to maximize the net energy flux.
   ii. Those coming from literature review and own estimates (hydropower, solar rooftop, geothermal, oceanic and waste): different potentials depending on low, medium, high assessments.
2. Endogenous given by the interaction with Land use submodule:
   i. Those coming from GIS analysis: solar PV on ground (Dupont et al., 2020). A global grid-cell methodology was adopted to assess the available global solar energy potential considering four constraints: land-use (suitability), solar irradiation, solar-to-electric technology, and net energy (EROI). Land-use requirements of capacity expansion are computed taking as reference the dynamic land-use efficiency (LUE, MW/km2), which refers to the power plant rated power capacity (W) per unit area of the total area of land occupied by the installation. LUE is endogenous for solar PV on ground depending on the evolution of the efficiency of solar PV panels and the latitude of the region. A full link with the Land and Water module is developed in a way that the Energy module demands a certain capacity of solar PV but Land use submodule check for land-use availability, eventually constraining its expansion. The method accounts for the fact that the potential may be constrained by a certain level of minimum EROI (EROImin).
   ii. Bioenergy coming from the own dynamics of the Land use submodule: liquid biofuels, biogases and solid bioenergy.
      1. For liquid biofuels and biogases, the limitation enters into play in the transformation chain in the transformation input shares (see section 4.5.2.2) which split between bio and fossil for liquids and gases. The user can set a target of supplying liquids and gases with bioenergy (having already subtracted the eventual production of synthetic fuels), which is conditioned by the availability from the Land use submodule.
      2. For solid bioenergy: land-use module includes a policy to limit the exploitation of forests, and computes a signal of availability of relative demand and supply of forestry products which enters into different points of the energy module: (1) drives substitution in final energy intensities, (2) affects the OPEX of solid bioenergy PROTRA technologies which affects utilization, and (3) affects the capacity expansion of solid bioenergy PROTRA technologies. Since most wood consumption in households is actually traditional biomass use, the policy which limits forest overexploitation is not affected by household demand.

[18] However, the kinetic level considered is very high and following other estimates such as (De Castro et al., 2011) would substantially decrease the wind potential from Dupont method.

It must also be highlighted that the potentials depending on the EROImin hold a significant uncertainty due to the difficulty to select a robust EROImin level to sustain complex societies. In fact, few estimates exist in the literature pointing to a 5-15:1 threshold (cf. e.g., (Brandt, 2017; de Castro & Capellán-Pérez, 2020; Hall et al., 2009). Literature is scarce and most is related with econometric analysis of past data for a selected number or regions. Moreover, a 'hard' EROI threshold to maintain a sustainable modern complex society cannot be established with precision *ex ante*, given that the reduced availability of discretionary energy as intermediary operations become less efficient is a gradual non-linear process with increasing and cascade consequences over time. This EROImin depends ultimately on social decisions, on how each society decides to allocate the available resources in case of decreasing EROI between investments and consumption considering also social inequalities. Hence, is more useful to think about ranges and increasing levels of risk. In this sense, between EROImin=5:1 and 15:1 there are differences of magnitude order in the regional potentials.

Dupont method assumes that the technical characteristics and performance factors of the technologies remain constant over time. This makes that for the same renewable energy inflow (wind speed, etc.) the EROItechn remains constant over time (efficiency in the numerator and energy investments per installed capacity in the denominator). Of course, EROItechn can vary in the future due to a number of reasons (technology improvement, increase in energy requirements for mining materials when ore decreases, curtailment due to VRES penetration, etc.; cf. discussion section 5.1 in (de Castro & Capellán-Pérez, 2020)).

4.5.2.6. VARIABILITY MANAGEMENT SUBMODULE

The Variability submodule consists of a set of multiple logistic regression models that emulate the results of the hourly energy model called EnergyPLAN based on a very large number of model runs.

Figure 24 summarizes the procedure behind this part of WILIAM. After the variables and values of EnergyPLAN are selected according to the importance of changes in the capacity factors of wind and solar power technologies, several input files are generated to simulate them with the energy model. The results are prepared for the regression analysis, in which a feature selection algorithm is applied to estimate the outputs and replicate the results of EnergyPLAN as close as possible. Finally, the analytical equations are introduced into the model, closing the loops to the capacity expansion of flexibility options and maximum full hours variable renewable technologies operate over the year. A detailed explanation of the method can be found in (Parrado-Hernando et al., 2023).

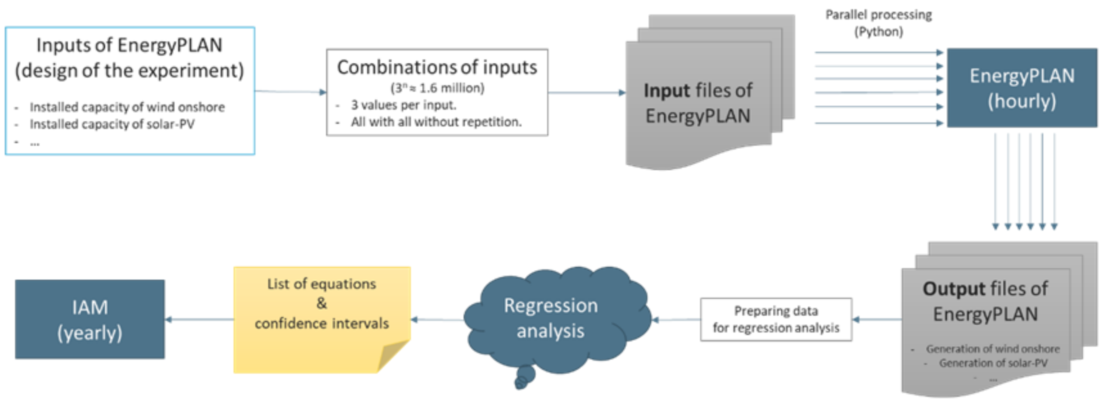


Figure 24: General schedule to integrate hourly features of EnergyPLAN into the IAM.

In WILIAM, the inputs are replicated as proxies (Figure 25). Then, the two main outputs, i.e., loss of capacity factor in variable renewable energies, have influence in the expansion of those technologies by limiting the potential in the allocation of the power system. A signal is useful also to further expand the capacities of the so-called flexibility options (electric boilers, heat pumps, pumped hydropower storage, flexible electrolysers, and demand-side management). Finally, the new installed capacities allow for additional production of energy. To remain consistency in the energy balances, the excess of electricity consumed by flexibility options is positively added to the electricity demand (TO, transformation output), while generations from flexibility options are negatively added (heat and hydrogen-based products). Consequently, the feedback of the regression models is double, to the capacity expansion and the transformation commodities.

Finally, the flexibility options here presented have been firstly modelled in WILIAM to follow the dynamics of these technologies based on a simple stock-flow structure, in line with the rest of technologies. A stock to represent the installed capacity, and two flows to explicitly simulate both the expansion and decommission of these special units. This has been the case of power-to-heat technologies (large heat pumps and electric boilers) and flexible electrolysers.

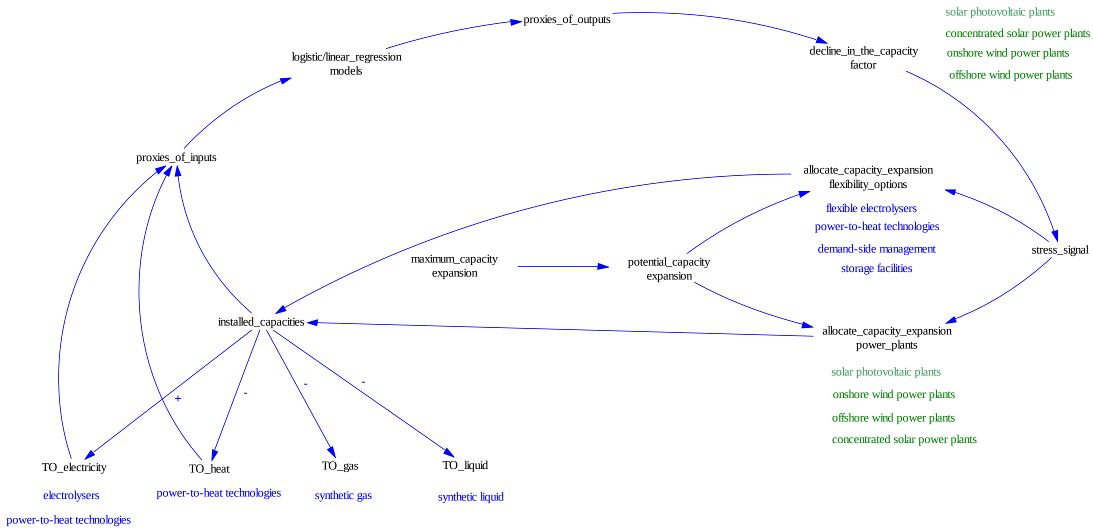


Figure 25: Causal loop of the energy Variability submodule.

4.5.2.7. GHG SUBMODULE

The Greenhouse gas emissions sub-module computes the emissions generated during the entire chain of energy transformations from the extraction of the main fossil resources (coal, oil and natural gas) to their final consumption in the different energy uses, including transformations, losses generated in their transport and distribution to end users, and their final consumption in both the economic sectors and households.

Emissions generated in the extraction, refining and distribution of the different energy flows are mainly fugitive emissions, due to losses or process needs. Emissions generated in the processes of transformation of input energies (TI) into output energies (TO) (electricity and heat) are combustion emissions. All of them are shown in Figure 26.

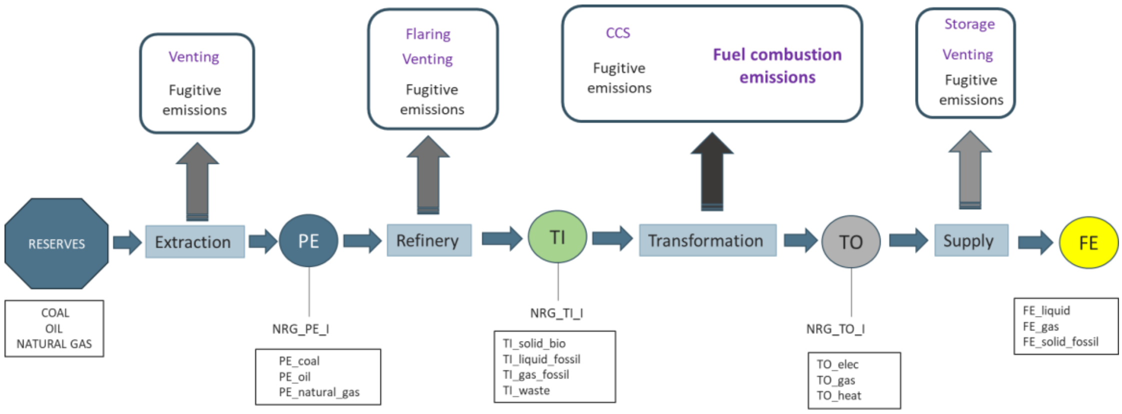


Figure 26: Reserves to final energy chain GHG emissions

Once the final energy reaches the users (households and sectors) the emissions generated depend on the process in which they are used. Those used in buildings (mainly heating) generate emissions from stationary combustion and therefore stationary type emission factors are used. Emissions generated in the transport use of households are mobile combustion emissions. While those emitted in the different productive sectors are both stationary in the different industrial processes of heat or electricity generation, and mobile in all transport systems of goods and materials. As you can see in the Figure 27.

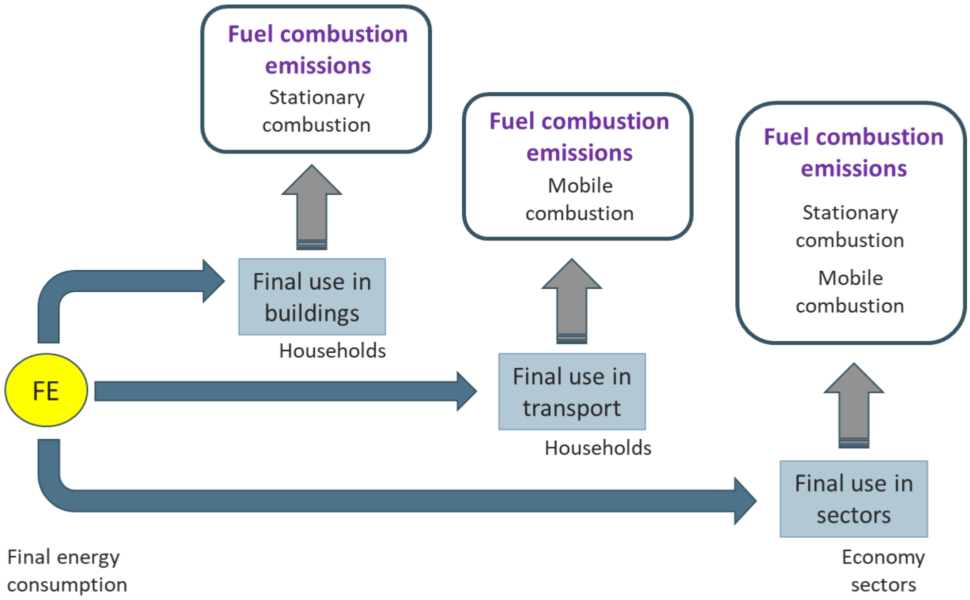


Figure 27: Final energy consumption GHG emissions

In addition to conventional energy sector emissions (Energy sector emissions), the model also computes emissions due to different manufacturing processes that do not involve the combustion of fuels such as the Industrial Processes and Product Use (IPPUs sectors). And also, emissions from the use and treatment of various municipal solid wastes (Waste emissions).

4.5.3. DATA SOURCES

A variety of sources was used to parametrize the Energy module:

- IEA extended energy balances (IEA, 2019)as primary data source for energy flows, parametrization, fuel shares etc.

- JRC IDEES capacity (Mantzos, et al., 2018) database for supplementary capacity data in European countries
- US Energy Information Administration as supplementary data source for energy transformation capacities (US EIA & EIA, 2020).
- International Renewable Energy Agency (IRENA, 2022)
- Building energy demand in China (Jiang, Yan, Hu, & Guo, 2016)
- Refinement to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories (IPCC, 2019).

4.5.4. IMPLEMENTED POLICIES

- Energy Efficiency Improvement
- Final energy substitution mechanism
- PROSUP capacity policy:
    - Battery storage expansion
    - Electrolyser capacity expansion
    - Electrolyzer capacity factor
    - Electrolyser utilization share for Hydrogen, synthetic gas or synthetic fuel production
    - PROSUP capacity expansion allocation priorities
    - PROSUP capacity expansion invest cost
    - Power2Heat expansion priorities (boilers, heat pumps)
- PROTRA Capacity expansion priorities (exogenous) & capacity expansion policy weight
- PROTRA capacity utilization allocation priorities (exogenous) & policy weight
- Bioenergy share of liquids and gas
- Availability of unmatured technologies selector (Electrolysers, floating offshore, CCS, oceanic)
- Exogenous RES potentials
- Wind and PV potential according to minimum EROI criterion
- Maximum on-roof PV installations
- Capacity investment cost development

4.5.5. KEY OUTPUTS

The most relevant outputs from the Energy module are as follows:

- Final energy consumption per capita
- Pkm by passenger transport mode
- Primary energy demand by commodity and region
- Related GHG emissions by economic sector and region
- Monetary investment and material and energy requirements (EROI) for energy capacities
- Land demand for PV installations

4.6. THE FINANCE MODULE

4.6.1. GENERAL DESCRIPTION

The Finance module of the WILIAM model aims to simulate the evolution of household wealth over time. In doing so, it affects their demand decisions and imposes financial constraints on the purchase of vehicles and housing. This module can simulate scenarios in all 35 WILIAM regions. For most European regions the module is disaggregated into 60 household types (as explained in the section on the Economy module). For those where no data were available, a single representative household has been used.

The three main inputs to the Finance module come from the Economy module. These three inputs are the gross savings of households, the interest rate on government debt by region (which is taken as the base interest rate) and the price of gross fixed capital formation of the real state sector. Starting from initial data on households' capital stock, households' financial assets and households' financial liabilities, households' net savings are accumulated, and these three variables are modified. Depending on the accumulated assets and liabilities, the property income that households receive or pay is calculated. This property income in turn serves to calculate households' disposable income, which is used to calculate gross savings that will change households' assets and liabilities, thus closing the feedback loop between the Finance and Economy modules.

On the other hand, there is a specific connection with transport so that the Finance module serves as a constraint to vehicle purchase. There are two key scenario parameters in this regard: the ratio of maximum annual loan payment over disposable income and the maximum years to repay a loan. These parameters, together with the interest rate on household liabilities, make it possible to calculate a maximum obtainable financing that restricts the purchase of cars if they cannot be financed with own resources and the indebtedness is above the maximum obtainable financing. The demand for cars also enters into the calculation of gross savings, closing a feedback loop between finance and transport. The representation of the relationships of the Finance module with other modules is shown in Figure 28.

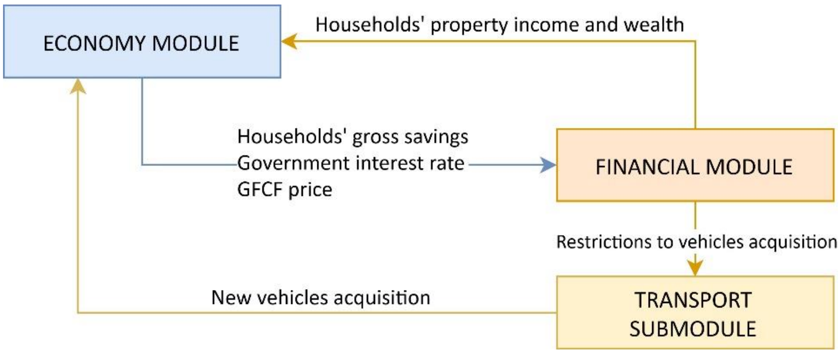


Figure 28: Financial module relationships with other modules and submodules.

4.6.2. METHODOLOGICAL APPROACH

The main equations of the model are represented below. The units of the model variables are $ for flows and $/year for stocks. First, the net wealth (NW) of a period is that of the previous period plus gross savings (S), plus the revaluation (rev) of real assets, less the depreciation (dep) of those assets. Given a certain ratio of liabilities to assets. (c), the net wealth and the amount of household assets and liabilities are as follows:

$$NW_t(R_i, H_j) = NW_{t-1}(R_i, H_j) + S\,(R_i, H_j) + rev\,(R_i, H_j) - dep\,\,(R_i, H_j)$$

Equation 113

$$Assets(R_i, H_j) = NW(R_i, H_j)/(1 - ratio\ liabilities\ to\ assets\ (R_i, H_j))$$

Equation 114

$$Liabilities\ (R_i, H_j) = Assets(R_i, H_j) - NW(R_i, H_j)$$

Equation 115

On the other hand, the stock of all assets and liabilities accumulated over time generates income and payments derived from ownership according to their interest rate.

$$Property\ income(R_i, H_j) = Assets\ (R_i, H_j) \times interest\ rate_{assets}(R_i, H_j)$$

Equation 116

$$Property\ outcome(R_i, H_j) = Liabilities\ (R_i, H_j) \times interest\ rate_{liabilities}(R_i, H_j)$$

Equation 117

The interest rate is calculated by adding the government interest rate (which is taken as the base) a mark-up calculated using historical data.

$$interest\ rate_{assets}(R_i, H_j) = interest\ rate_{gov\ debt}(R_i, H_j) + mark\ up_{assets}(R_i, H_j)$$

Equation 118

$$interest\ rate_{liabilities}(R_i, H_j) = interest\ rate_{gov\ debt}(R_i, H_j) + mark\ up_{liabilities}(R_i, H_j)$$

Equation 119

It is a maximum level of indebtedness that serves to restrict the demand for vehicles in the Transport submodule. First, the maximum payment that households can make annually for interest payments (mlp) and debt repayments is calculated (this maximum payment is taken to be a certain percentage, p, of disposable income). Then, using this maximum payment, the interest rate on the debt and the maximum number of years to repay the debt (y), the households' maximum financing obtainable (mfo) is calculated. Once this maximum amount is reached, they can only buy vehicles with their disposable income and not by borrowing.

$$mlp(R_i, H_j) = YD(R_i, H_j) \times p(R_i, H_j)$$

Equation 120

$$mfo(R_i, H_j) = \frac{mlp(R_i, H_j) \times \left(\left(1 + interest\ rate_{liabilities}(R_i, H_j)\right)^{y(R_i,H_j)} - 1\right)}{\left(1 + interest\ rate_{liabilities}(R_i, H_j)\right)^{y(R_i,H_j)} \times interest\ rate_{liabilities}(R_i, H_j)}$$

Equation 121

4.6.3. DATA SOURCES

4.6.3.1. EU27-COUNTRIES

Data for EU countries are from Eurostat: https://ec.europa.eu/eurostat/en/web/main/data/database

4.6.3.2. NON EU27-COUNTRIES

For regions outside the European Union, an extrapolation has been made based on population and financial values of households in low-income countries in Europe.

4.6.4. IMPLEMENTED POLICIES

a. Base interest rate (interest rate of government debt).
b. Ratio of maximum annual loan payment over disposable income.
c. Maximum years to repay a loan.

4.6.5. KEY OUTPUTS

a. Households' capital stock

b. Households' financial assets
c. Households' financial liabilities
d. Households' property income paid
e. Households' property income received
f. Households' net wealth

4.7. THE LAND AND WATER MODULE

4.7.1. GENERAL DESCRIPTION

The Land and Water module is one of the eight modules belonging to the WILIAM model. It is structured according to several submodules as shown in Figure 29, which reflects the relationships of Land and Water submodules with the rest of the model and the allocates related to this module. The Land and Water module is working in 9 global regions (EU27, UK, CHINA, EASOC, INDIA, LATAM, RUSSIA, USMCA, and LROW).

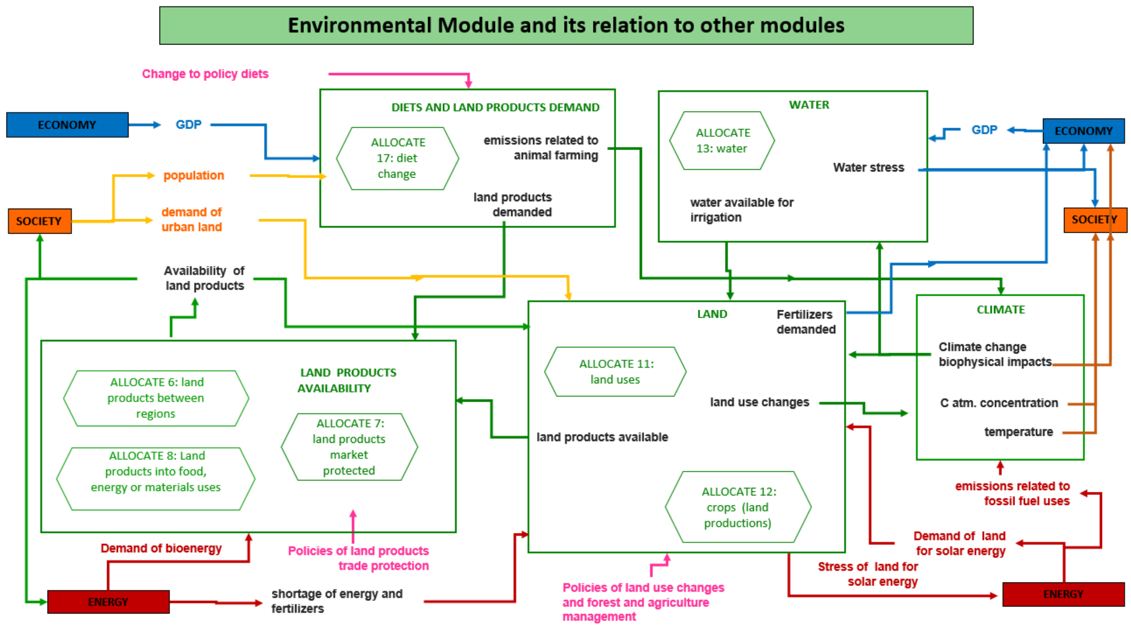


Figure 29: Relation of Land and Water module with the rest of the WILIAM model.

Figure 29 shows the following interactions between the Land and Water with the rest of WILIAM model:

- With the Economy module: receive the Gross Domestic Product per capita (GDPpc) to calculate the demand for land products for food, they also receive the production of the sectors (output real) to calculate the demand for wood for industry and water and send the demand for fertilizer and the indicators of water stress.
- With the Society and Demography modules: takes population and demand of land for urbanization and provides shortage of land products and diets.
- With the Energy module: takes primary energy by commodity and the emissions from energy and material use and the land demand for solar energy and gives Energy module the shortage of land products availability and the shortage of solar land.
- The Land and Water module communicates with other modules through Land products availability submodule to which sends the availability of land products and receives the shortage of energy, fertilizers, and land products.

4.7.2. SUBMODULES

The Land and Water module has been split into several submodules which are structured as shown in Figure 30, and the main tasks of each submodule are as follows:

- The Diets and land products demand submodule calculates diets according to GDP and policies (Flexitarian, Willet, Baseline, 100% plant based, and 50% plant-based diets), the land products demanded for food by using an Agro-food transformation matrix, and the land products demanded for energy and industry that will be sent to the Land products availability submodule to calculate the shortage of availability of land products.
- The five submodules related to Land (Land uses, Croplands, Yields, Forests, and Grasslands) calculate the availability of land products from forests and agricultural lands.
- The Water submodule calculates the demand of water for several uses and the availability of water for irrigation as a function of temperature and carbon concentration in the atmosphere and determines the availability of water for irrigation and other uses.
- Biophysical impacts of climate change are calculated in the Emissions submodule such as carbon dioxide, methane, and nitrous oxide emissions, and are sent to the Land submodules to calculate agricultural production.
- Finally, in the Land products availability submodule, the supply and demand of land products are compared along with others for all uses (food, energy, industry) and the availability of resources, the possible shortage, is sent back to Land submodules.

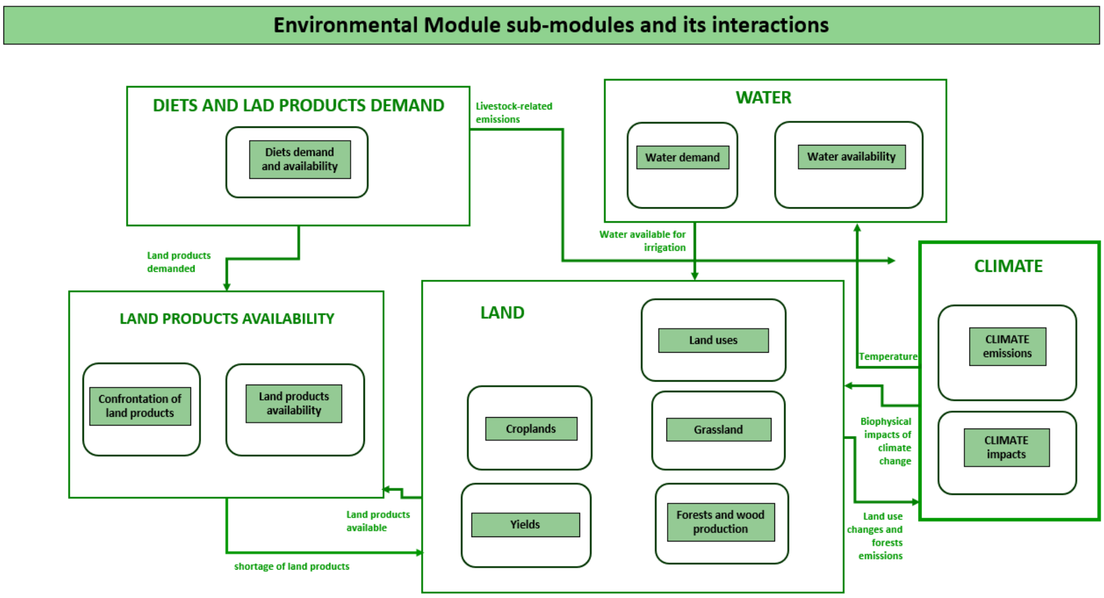


Figure 30: Submodules of the Land and Water module.

4.7.2.1. DIETS AND LAND PRODUCTS DEMAND SUBMODULE

The demand for food depends on the diets and is defined in terms of products directly consumed by consumers that are converted to land products demanded by the whole food industry and farm transforming system (see Figure 31). The demand for food is calculated based on data of GDPpc, population and diet. Diets are calculated based on dietary patterns observed from historical data and a target diet policy. The resulting estimation of the diet would be the combination of the one given by GDPpc and the policy one. The relation between the diet products and the land products depends on the Agro-food transformation matrix that states the transformation of the whole farming and food industry system. The amount of Land products demands for food that depends on land uses is calculated by subtracting those sources of food that do not compete directly with cropland use (such as fish, dairy, and meat directly from grasslands). In this submodule, the land products demanded for energy and industry are calculated by using primary energy by commodity from Energy module and output real for forestry sector from Economy module.

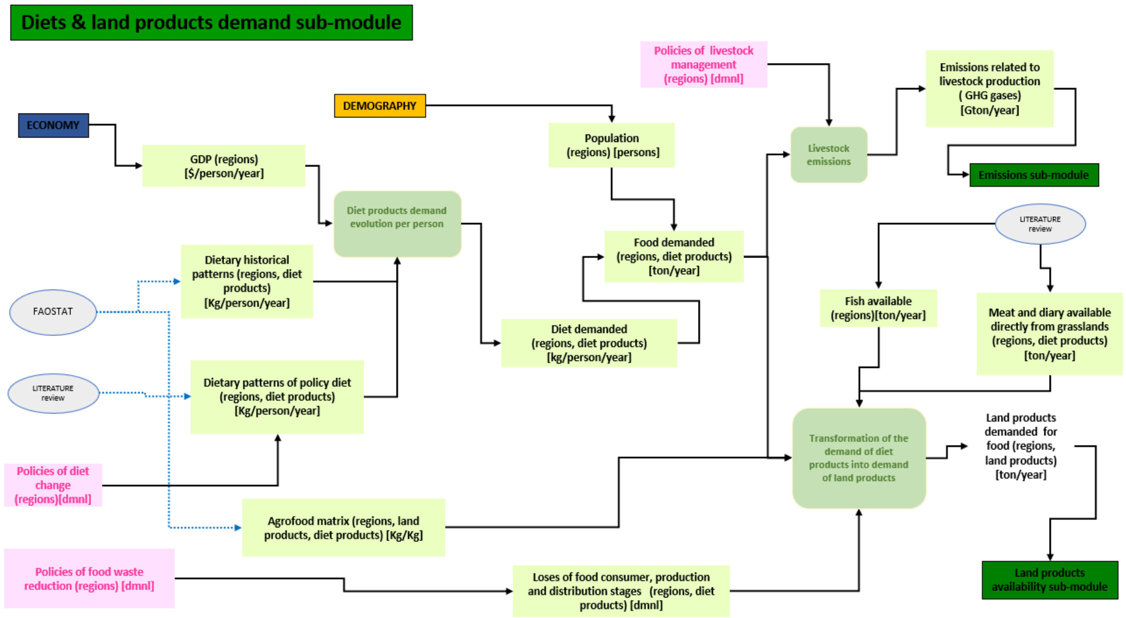


Figure 31: Diets and Land Products Demand submodule structure.

4.7.2.2. LAND USES SUBMODULE

This submodule oversees allocating the land among several uses (rainfed and irrigated croplands, managed forests, primary forests, and tree plantations, grasslands, urban lands, solar lands for energy, and other uses) by taking the demands for land use changes coming from the population demand for food, for urban lands, for solar energy or for biomass. It uses signals such as the shortage of land products and energies that allow it to calculate the land stress for several uses and calculates the loss of agricultural land due to sea level rise. Figure 32 shows the main elements of the Land uses submodule.

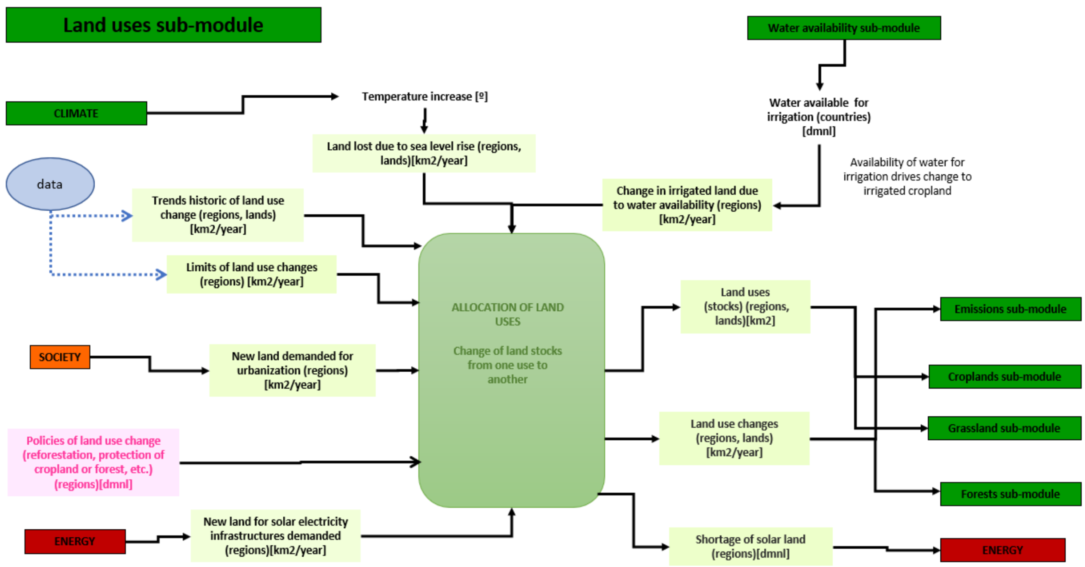


Figure 32: Land uses submodule structure.

4.7.2.3. CROPLANDS AND YIELDS SUBMODULES

The Croplands and Yields submodules (Figure 33 and Figure 34) are in charge of the agricultural production. Croplands manages the cropland irrigated and rainfed separately and crop yields are calculated in the Yields submodule for 10 types of crops (corn, rise, cereals, tubers, soybeans, pulses and nuts, oil crops, sugar crops, fruits and vegetables, and other crops). These crop areas are taken from historical data and may vary with the demand and the shortage that may occur at any given moment.

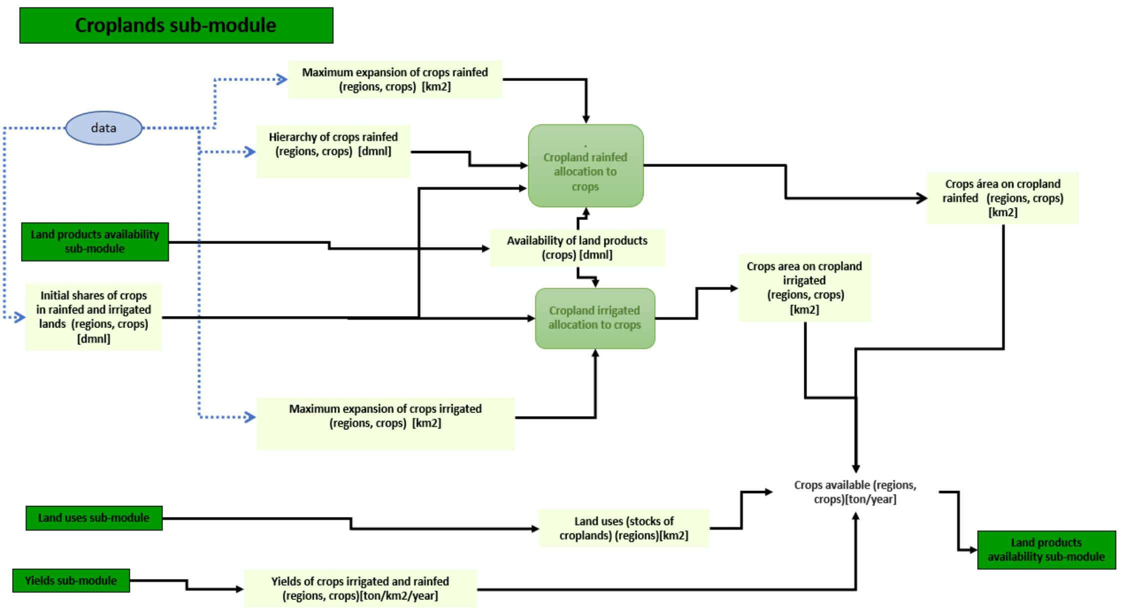


Figure 33: Croplands submodule structure.

Agricultural yields are conditioned by several things: the effects of climate change (that are calculated in by Climate module), the expected improvement of yields, the possible shortage of fertilizers and liquid fuels and the change of agricultural methods towards regenerative methods to respond to fertilizer shortage or to policies of biodiversity protection.

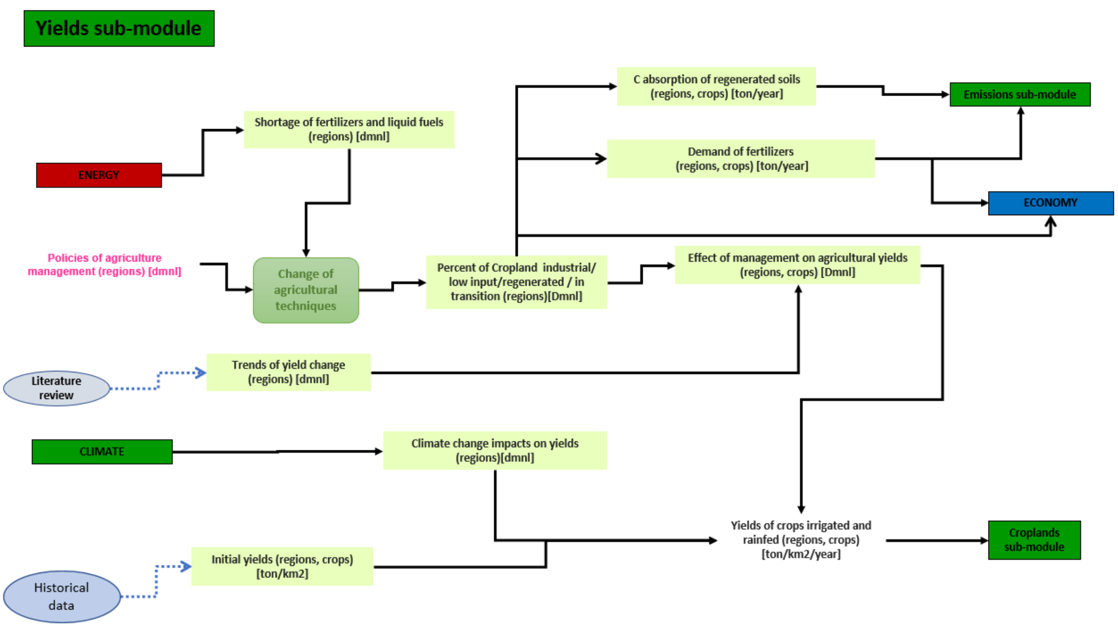


Figure 34: Yields submodule structure.

4.7.2.4. GRASSLANDS SUBMODULE

The Grasslands submodule (see Figure 35) calculates the absorption of carbon captured in grasslands due to change in management. It incorporates the possibility of improved capture and increased yields by regenerative management.

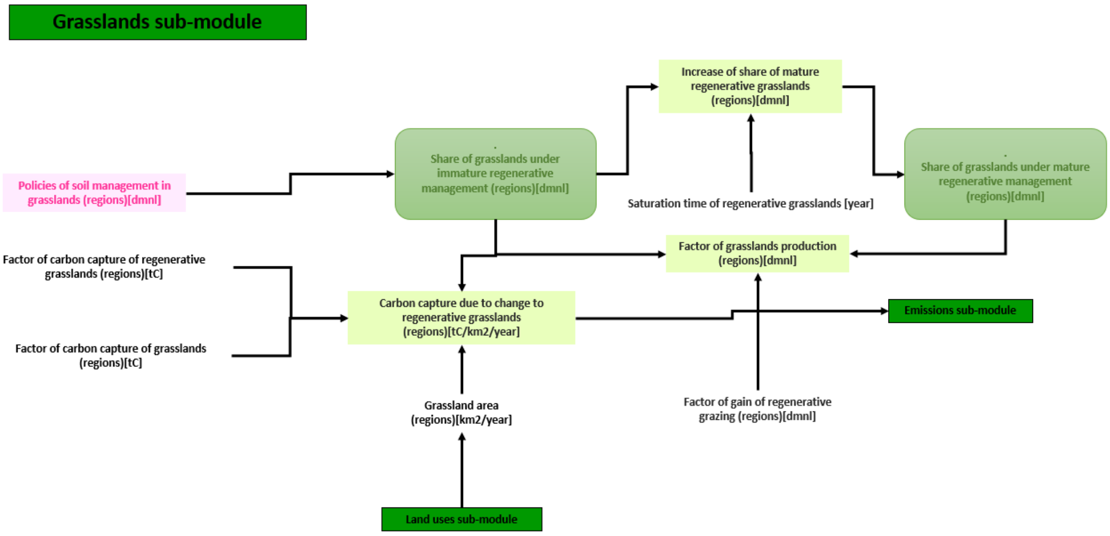


Figure 35: Grasslands submodule structure.

4.7.2.5. FORESTS SUBMODULE

The Forests submodule (Figure 36) accounts for carbon flows in forests in order to know the carbon stock that is contained in the forests and relate it to the wood extracted and the land lost by deforestation. This will allow us to obtain a potential for sustainable timber extraction from forests.

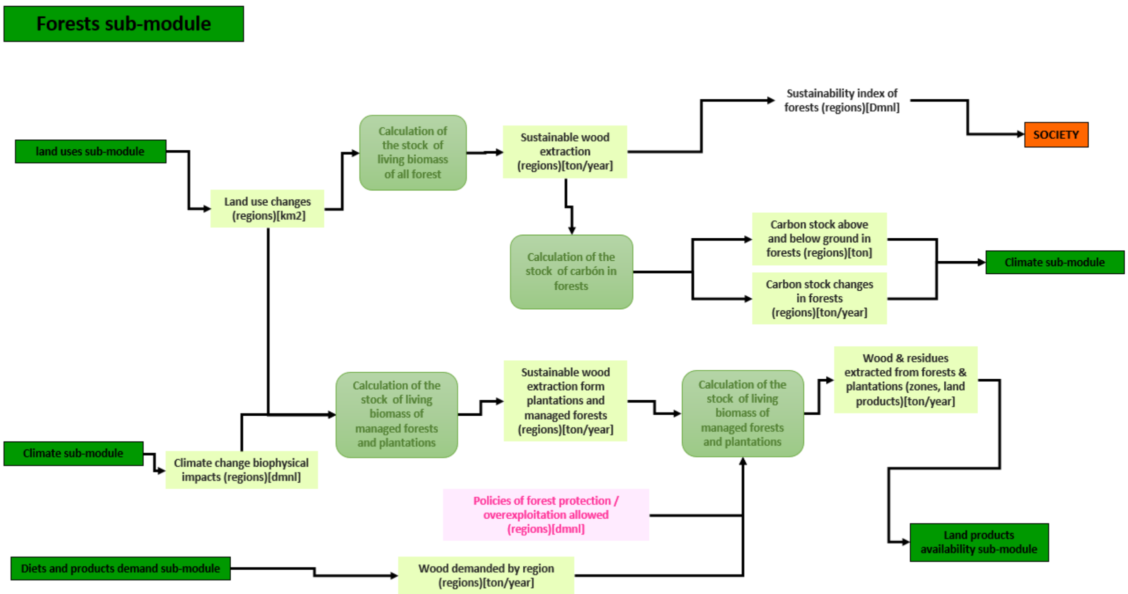


Figure 36: Forests submodule structure.

4.7.2.6. LAND PRODUCTS AVAILABILITY SUBMODULE

The Land products availability submodule (see Figure 37 and Figure 38) is composed of two parts, the first one is related to the land products availability where Land products available to global market and protected from global market are estimated, and the second one is related to the confrontation of land products where the supply and demand of land products are compared along with others for all uses, food, energy, and industry, and the availability of resources, then the possible shortages are sent back to Land submodules and Society and Energy modules.

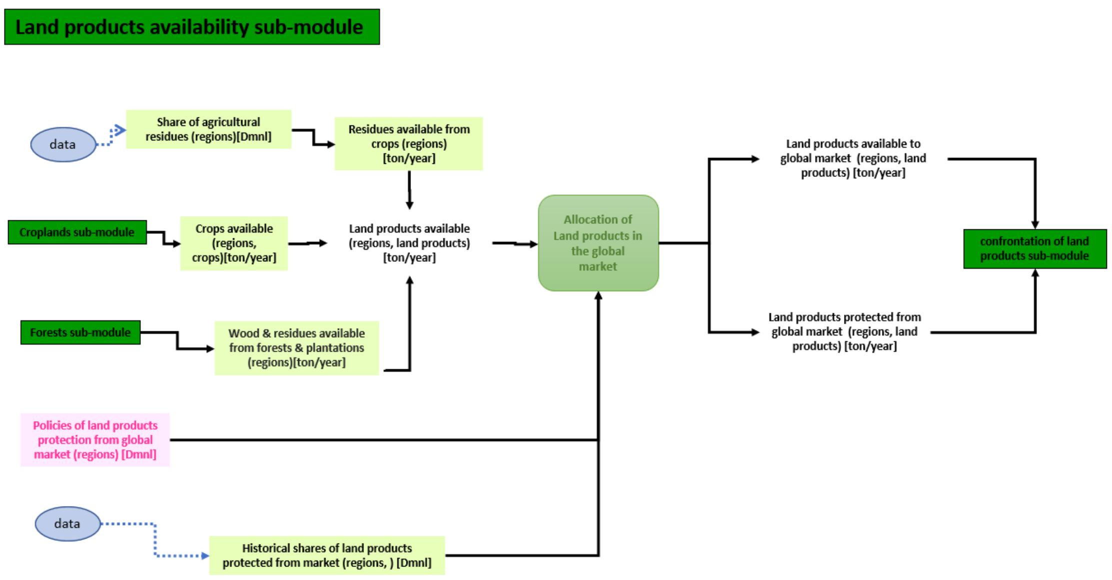


Figure 37: Land products availability submodule structure.

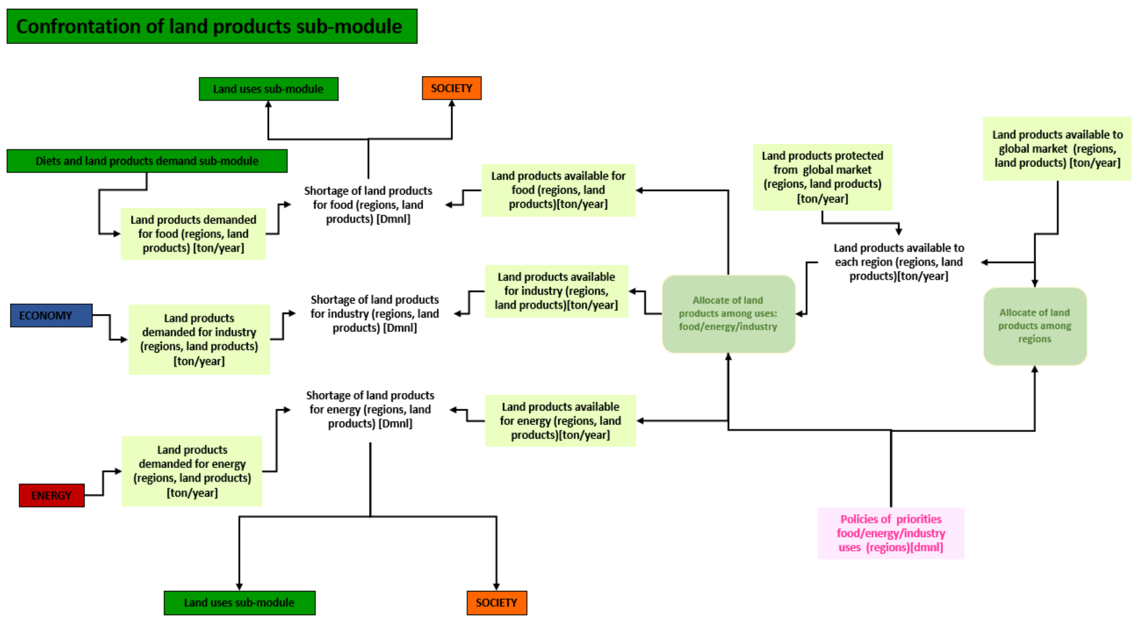


Figure 38: Confrontation of land products submodule structure.

4.7.2.7. EMISSIONS SUBMODULE

The Emissions submodule calculates endogenously agriculture (fertilizers, rice crops and livestock) and LULUCF (Land Use, Land-Use Change and Forestry) emissions, i.e., emissions from land use changes and uptake by forests and other land with vegetation (see Figure 39). In particular, CO_2 (carbon dioxide), CH4 (methane), and N2O (nitrous oxide) emissions. In the diagram, it is possible also to detect the links with the other modules and other parts of the Land module.

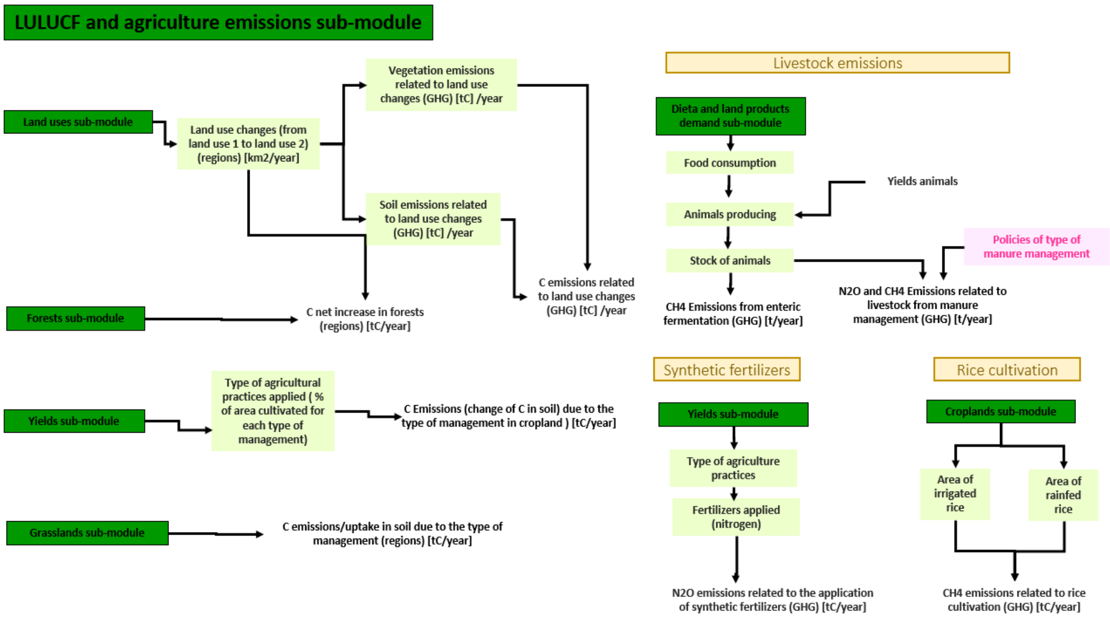


Figure 39: Emissions submodule structure.

4.7.2.8. WATER SUBMODULE

The Water submodule consists of computing the water demand (Figure 40) and water availability (Figure 41) in the 9 regions. Water demand values were computing for green, blue, and blue water for households. Precipitation/Evapotranspiration (P/E) values were computed using different equations for each 9 regions, depending on the radiative forcing, to calculate the water availability. The Water demand for agriculture is also dependent on the P/E due to, for example, higher need of water with increasing evapotranspiration. The Water submodule has links with the Economy, Demography, Energy modules, and the Land uses submodule, with special emphasis on the hydropower potential, water demand for households and irrigated land.

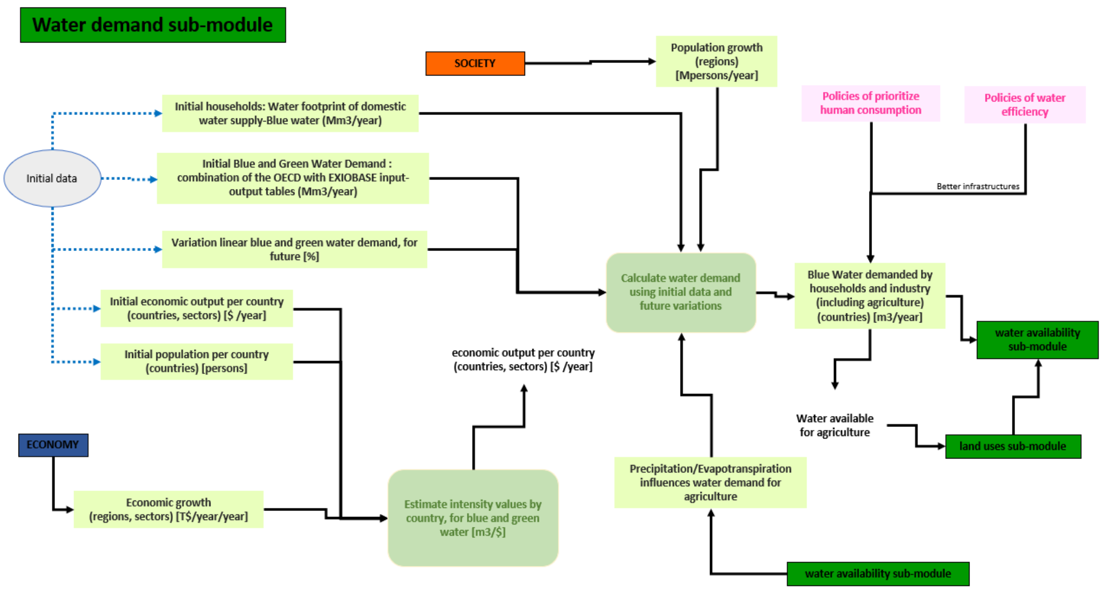


Figure 40: Water demand submodule structure.

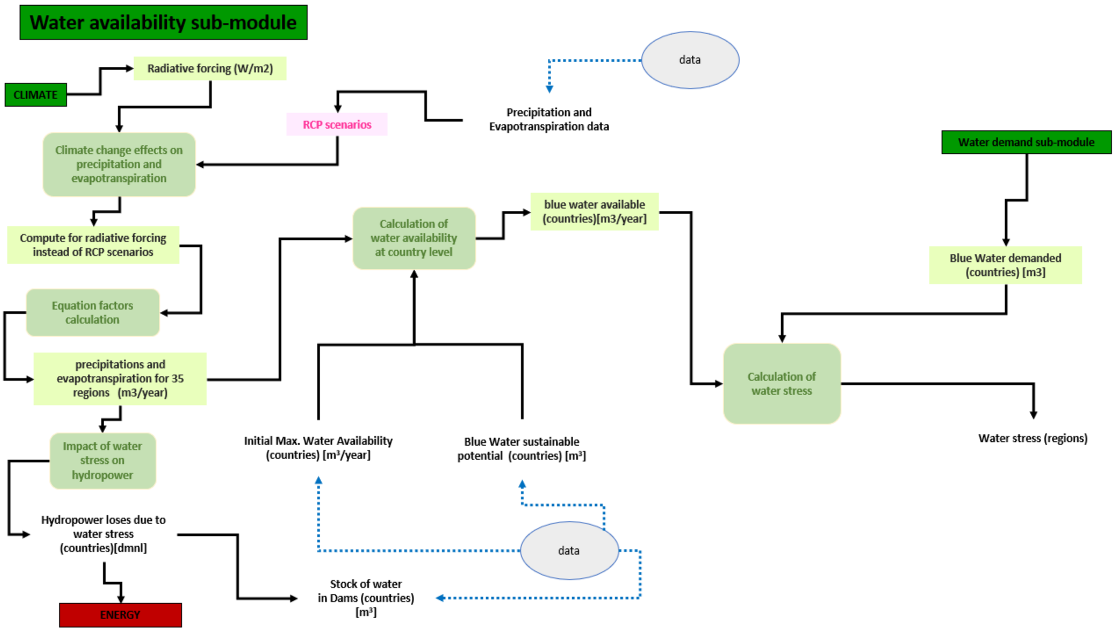


Figure 41: Water availability submodule structure.

4.7.3. METHODOLOGICAL APPROACH

The WILIAM model is aggregated into regions based on economic criteria related to the Input/Output (I/O) framework. This disaggregation is useful to manage everything related to the economy and to the demands of some physical goods such as energy, food, and water. However, when dealing with the resources extracted from the land, these economic regions are not the most suitable. Most of the biophysical parameters that are of interest for the land model, such as crop yields, forest growth, water availability or irradiance (for solar energy), depend on climatic factors rather than on the political boundaries of countries. The Land and Water module has developed a mixed methodology that would solve these difficulties without turning the Land use submodule into a grid-based model (which is something out of the reach of the project). This is based on a double disaggregation: for some aspects it will be based on WILIAM's current economic regions and, for others, based on the newly defined climate zones.

- WILIAM's regions R_i with i = 1...M (M=9 LOCOMOTION regions).
- Food F_j with j = 1...N (N = 14 Food categories).
- Land uses U_n with n = 1...L (L = 12 Land use categories).
- Land products Lp_k with k = 1...D (D = 13 Land products categories).

The main equations used in the Land and Water module are listed as follows:

4.7.3.1. FOR DIETS AND LAND PRODUCTS DEMAND SUBMODULE

-The Diet demanded [$DD(R_i, F_j)$, kg/(year*person)] (Equation 122), is calculated based on historical data from 14 diets patterns (Cereals; Tubers; Pulses, Legumes, and Nuts; Fruits and Vegetables; Fats Vegetal; Fats Animal; Dairy; Eggs; Meat Ruminants; Meat Monogastric; Fish; Sugars; Beverages; Stimulants) by GDPpc for 9 regions [$DG(R_i, F_j)$, kg/(year*person] and diets by policies [$DP(R_i, F_j)$, kg/(year*person] related to the desired sustainable diet to be obtained when policies of diet change are applied (options are: flexitarian, willet, baseline, 100% plant based, and 50% plant based). The dynamic change between present and desired diet is driven by the variable change to policy diets [$SPC(R_i)$, dmnl] that starts with a value 0 and reaches 1 when the full diet change is achieved:

$$DD(R_i, F_j) = DP(R_i, F_j) \cdot SPC(R_i) + DG(R_i, F_j) \cdot (1 - SPC(R_i))$$

Equation 122

-Land products demanded for food [$LPDF(R_i, Lp_k)$, ton/year] is calculated multiplying the food demanded by households [$FH(R_i, F_j)$, ton/year] = $DD(R_i, F_j) * pop(R_i)$] by an Agro-food transformation matrix [$AM(F_j, Lp_k)$, dmnl] that related food products and land products (Equation 123):

$$LPDF(R_i, Lp_k) = FH(R_i, F_j) \cdot AM(F_j, Lp_k)$$

Equation 123

-The Land products demanded for energy [$LPDE(R_i, Lp_k)$, ton/year] (Equation 124) is calculated based on data of agriculture products demanded [$APDE(R_i)$, ton/year] that comes from the Energy module via primary energy by commodity variable and the percentage of each land product used for energy use [$LPP(\ Lp_k)$, dmnl] (we assume it constant):

$$LPDE(R_i, Lp_j) = APDE(R_i) \cdot LPP(\ Lp_j)$$

Equation 124

-Land products demanded for industry variable [$LPDI(R_i, Lp_k)$, ton/year] is calculated based on data of output real of forestry sector [$OR(R_i)$, Mdollars 2015/year] by region and sectors (that come from economy model and historical data of intensities of roundwood [$IRW(R_i)$, ton/Mdollars 2015] and wood residues [$IWR(R_i)$, ton/Mdollars 2015] for industry (Equation 125):

$$LPDI(R_i, Lp_j) = OR(R_i) \cdot (IRW(R_i) + IWR(R_i))$$

Equation 125

4.7.3.2. FOR LAND SUBMODULE

-The matrix of LUC (Land Use Change) demands equation is expressed as follows (Equation 126):

$$Matrix\ of\ LUC\ Demands\ (R_i, L_n, L_m) = LUC\ Demands\ (R_i, L_n) \cdot Share\ of\ LUC\ from\ Others\ (R_i, L_n, L_m)$$

Equation 126

Where the variable $LUC\ Demands\ (R_i, L_n)$ is depending on two variables, the first one is trends of LUC (R_i, L_n), that is calculated based on historical data from 12 types of Lands (Cropland Rainfed; Cropland Irrigated; Forest Managed; Forest Primary; Forest Plantations; Shrubland; Grassland; Wetland; Urban Land; Solar Land; Snow Ice Waterbodies; Other Land) and for 9 regions, the second one is LUC driven by demands (R_i, L_n) that includes the increment of population (Demography module), the increment of land for solar use (Energy module), the shortage of rainfed and irrigated crops (Land products availability submodule), the shortage of wood (Land products availability submodule), and the reforestation policies (Land use submodule). The variable $Share\ of\ LUC\ from\ Others\ (R_i, L_n, L_m)$ is calculated by using some historical data from 12 types of Lands (L_j) and for 9 regions (R_i).

Then the land use changes of matrix of LUC demands might not be fulfilled if policies of land use protection of limits to land use conversion are activated. The $Matrix\ of\ LUC\ Demands\ (R_i, L_n, L_m)$ is the one that stores the achievable changes and it is collapsed into a $vector\ LUC\ (R_i, L_n)$ by adding the changes that give to each use and subtracting the ones that demand from it (Equation 127):

$$vector\ LUC\ (R_i, L_m) = \sum_n matrix\ LUC(R_i, L_n, L_m) - \sum_n matrix\ LUC(R_i, L_m, L_n)$$

Equation 127

4.7.3.3. FOR EMISSIONS SUBMODULE

-General equation estimating the total change transfer of carbon (Equation 128):

$$\begin{aligned}\Delta C_{(region)} &= \left(A\ (t2) \cdot Cdensity_{t2,l2(region)} - A\ (t1) \cdot Cdensity_{t1,l1(region)}\right) \\ &= A \cdot Cdens._{t2,l2(region)} - A \cdot Cdens._{t1,l1(region)} \\ &= A \cdot Factor\ emission_{l1-l2,region}(tC)\end{aligned}$$

Equation 128

Where $C\ density_{t2,l2(region)}$ is the carbon density (tC/Mha) in the year "t" and the land use "X" in zone "region", being "1" the prior land use, and "2" the new land use, A is the area of land that is being changed (Mha).

After applying the previous equation, afterwards is necessary to allocate the change of carbon across time. For these two approaches has been applied:

- Instantaneous emissions (emitted in the same year)
- Applying an exponential delay function of first order (asymptotic function) considering the equilibrium period (time of carbon to reach the equilibrium to a new state). In this case the form of the equation is the following (Equation 129):

$$C_{(t)} = (C_0 + a\ [1 - e^{-bx}]) \rightarrow \left(C_{stock\ 0} + a\left[1 - e^{-\frac{1}{\tau}*t}\right]\right)$$

Equation 129

being τ the time equilibrium period.

-Equation to represent emissions from a particular activity, being in these case agricultural emissions (Equation 130):

$$Emission = EF \cdot A$$

Equation 130

Where EF is the emission factor, and A is the activity parameter, which depends on the type of activity:

- "Number and type of animals" for livestock N_2O and CH_4 emissions (from manure management and enteric fermentation)
- $“rice\ paddy\ annual\ harvested\ area"\ (ha^2)$ for estimation of CH_4 emissions rice cultivation emissions
- $"N\ application"\ (kgN/year)$ for estimation of N_2O emissions from the application of synthetic fertilizers by farmers.

4.7.3.4. FOR WATER SUBMODULE

-The water stress [$WS(R_i)$, m3] (Equation 131) is calculated by dividing the total demand of blue water from households and by region [$BWD(R_i)$, m3] = $BWDP(R_i) * pop(R_i)$] by the availability of water by region [$WA(R_i)$, m3] (Equation 132):

$$WS(R_i) = BWD(R_i)/WA(R_i)$$

Equation 131

where,

$$WA(R_i) = [ITRW(R_i) \cdot MEWC \cdot PEP(R_i)]/IPEP(R_i)$$

Equation 132

With, $ITRW(R_i)$ is the Initial Total Renewable Water by region; $MEWC$ is the Maximum Exploitation Water Coefficient; $PEP(R_i)$ is the Precipitation Evapotranspiration Projections by region; and $IPEP(R_i)$ is the initial precipitation evapotranspiration by region.

4.7.4. DATA SOURCES

The most relevant data of Land and water module have been collected from FAO STAT and OECD databases, and most searched data are from 2005 to 2020.

4.7.5. IMPLEMENTED POLICIES

The Land and Water module contains several policies that affect the main outputs of the module, and they are listed as follows:

- Change to diets according to policies (Diets and land products demand submodule).
- Afforestation policy (Land uses submodule).
- Land protection policy (Land uses submodule).
- Priorities of land use change policy (Land uses submodule).
- Urban land density policy (Land uses submodule).
- Solar land from others policy (Land uses submodule).
- From traditional to industrial agriculture policy (Croplands & yields submodules).
- Change to regenerative agriculture policy (Croplands and yields submodules).
- Effect oil and gas on agriculture policy (Croplands and yields submodules).
- Land products global pool policy (Land products availability submodule).
- Crops for energy policy (Land products availability submodule).
- Wood for energy policy (Land products availability submodule).
- Forest overexploitation policy (Forests submodule).
- Soil management in grasslands policy (Grasslands submodule).
- Prioritize human consumption policy (Water submodule).
- Water efficiency policy (Water submodule).
- Solar land management policy (Emissions submodule).

4.7.6. KEY OUTPUTS

The most relevant outputs of Land and Water module are as follows:

For Diets and land products demand submodule:

- Food availability
- Food demand
- Land products demanded for food

For Land uses submodule:

- Land use area by region
- Land use changes demanded

- New land for solar demanded
- Urban land density
- Cropland loss due to sea level rise by region

For Croplands and yields submodule:

- Land products available from croplands
- Area of rainfed crops estimated
- Area of irrigated crops estimated
- Yields of irrigated crops
- Yields of rainfed crops

For Forests submodule:

- All forests stock
- Forestry sustainability index global
- Forest carbon total stock
- Forest carbon dioxide total flow global
- Roundwood extracted
- Managed forest and plantations stock

For Grasslands submodule:

- Carbon capture due to change to regenerative grasslands
- Share of grasslands under regenerative grasslands

For Water submodules:

- Water available for agriculture
- Total demand water by region
- Water stress by region

From Emissions submodule:

- Agriculture emissions:
    - N2O (nitrous oxide) emissions from synthetic fertilizers application
    - CH4 (methane) emissions from rice crops
    - N2O (nitrous oxide) and CH4 (methane) emissions from livestock
- LULUCF (Land Use, Land-Use Change and Forestry) emissions, i.e., emissions from land use and land use changes, including cropland and grassland management practices.

4.8. THE CLIMATE MODULE

4.8.1. GENERAL DESCRIPTION

The Climate module is closely linked to the Land and Water modules, which is structured according to the submodules shown in Figure 29 in the previous section. Below, in the zoom of the previously mentioned figure, it is possible to see the link of the Climate module with the rest of the modules and submodules of WILIAM, including also the Society, Economy and Energy modules.

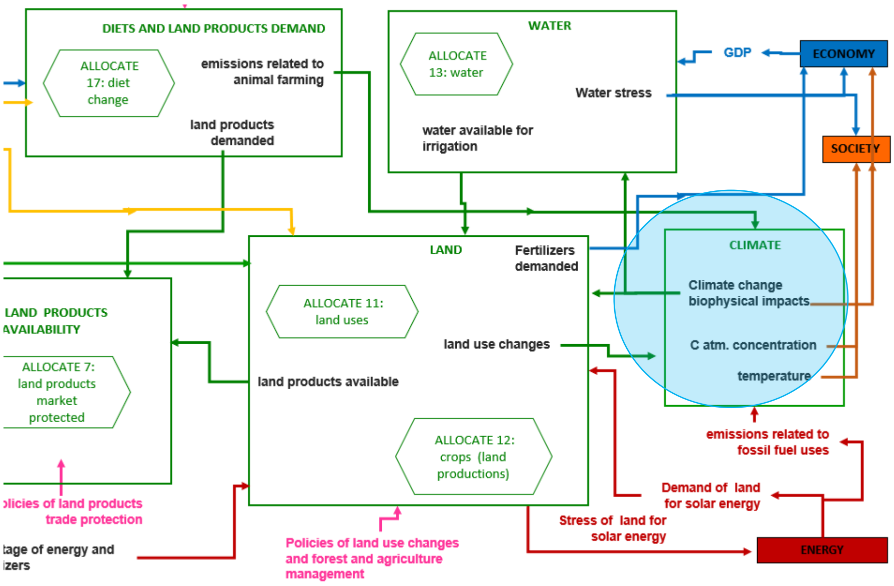


Figure 42: Interactions of the climate module with the rest of the WILIAM model.

The interactions of the Climate module with the rest of the model (Figure 42) are the following:

- With *Energy:* The Climate module takes the emissions related to fossil fuel uses (transport, energy consumption, etc.) calculated in the Energy submodule.
- With *Land:* The Climate module takes the emissions related to agriculture, land use change, and forestry emissions. From the submodule "LULUCF and agriculture emissions submodule".
- Climate change impacts: it provides with the change in main climate variables, such as global and regional mean temperature changes. This allows to estimate climate change impacts in several modules through damage functions. In particular with the *Economy, Water, Land and Society* modules. In this case, the direction in this case is "FROM" the Climate module to the rest of the modules affected.

The Climate module in WILIAM can be subdivided in the following parts (Figure 43):

- Energy, agriculture and LULUCF emissions. The emissions coming from energy, economy and land modules enter in the climate module as inputs. The different emissions calculated endogenously in WILIAM can be seen in Figure 43.
  - CO_2 emissions due to land use changes
  - CO_2 emissions/uptake due to change in the type of management of cropland and in grassland.
  - CO_2 emissions/uptake derived from forestry practices.
  - CO_2, N_2O, and CH_4 emissions from energy consumptions (including transport).
  - CH_4 emissions related to agriculture activities (rice cultivation, livestock: manure management and enteric fermentation)
  - N_2O emissions related to agriculture activities (livestock: manure management, and synthetic fertilizers)

  The greenhouse (GHG) emissions not calculated endogenously are consistent with the RCP scenarios (Representative Concentration Pathways).
- Climate variables modelling. The main subparts including in this part of the Climate module are based in C-ROADS simple climate model, which is an emulator of complex climate models (Capellán-Pérez et al., 2020; Fiddaman et al., 2018), and which has been adapted and integrated into WILIAM structure and requirements. This includes:

  - GHG cycles. The cycle of each GHG is modelled separately, as well as the interactions between cycles. The cycles included in WILIAM are the carbon cycle, CH_4 (methane), N_2O (nitrous oxide), PFCs (perfluorocarbons), SF_6 (sulphur hexafluoride), and HFCs (hydrofluorocarbons).
  - The modelling of climate variables. This includes the modelling of carbon (C) and other GHG gases concentration, to calculate the total radiative forcing of each gas. The main outputs of the climate module are the total radiative forcing, the mean global temperature change, the seal level rise and the ocean acidification.
- Climate change impacts modelling, which includes:
  - Regionalisation of climate variables at each climate zone, to model impacts considering the heterogeneity of climate in different areas of the Earth.
  - Modelling of impacts on forests (on Net Primary Productivity), on water availability, on the land area occupied due to sea level rise, and in crop yields. This part links the Climate module with the “Land and Water” modules.

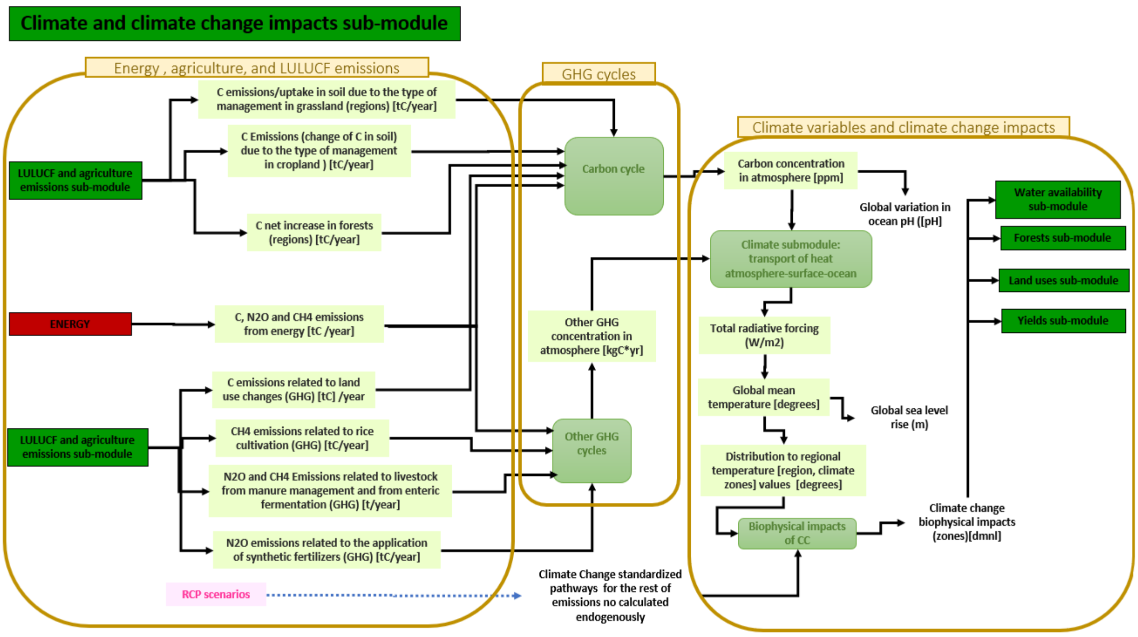


Figure 43: Structure and main parts of the Climate module.

Finally, five tipping points are also assessed in the Climate Module, with increasing probabilities depending on the temperature increase in future climate. Specifically, the Atlantic Meridional Overturning Circulation (AMOC) weakening is a tipping point that can be activated and has impacts in regional temperature changes in the northern hemisphere, with low probability. The other modelled tipping points are the Melt of the Greenland Ice Sheet, the Disintegration of the West Antarctic Ice Sheet, the Dieback of the Amazon Rainforest and the Shift to a more persistent El Niño Southern Oscillation. On the other hand, the permafrost tipping points whose effects will feed directly into the carbon and methane cycles is by default deactivated due to its uncertainty.

4.8.2. METHODOLOGICAL APPROACH

Although the emissions entering into the Climate module of WILIAM are at regional scale, it is necessary to mention that the climate module is at global scale. The GHG cycles, the radiative forcing, and the energy balance that allow to calculate the impact of the radiative forcing in the mean temperature change, need to be modelled at global level. In this sense, the main climate variables obtained are at global level. However, when dealing with the modelling of climate change impacts, it was detected the need for downscaling (or distribution among the regions) the global climate variables obtained, and specifically the temperature change, to represent better the different climates and therefore unequal distribution of the impacts of climate change among global regions.

The main equations used in the Climate module, and specifically, for obtaining key climate variables, are the following:

1. The main and general equation for introducing the emissions and modelling GHG cycles is the following:

$$\mathrm{GHG_{conc}}(Mtons) = GHGemiss_{natural} + GHGemiss_{anthropogenic} - GHG_{uptake} = GHGemiss_{natural} + GHGemiss_{anthropogenic} - GHG_{conc} \times \frac{1}{GHG_{time}}$$

Equation 133

Where $\mathrm{GHG_{conc}}(Mtons)$ is the GHG concentration in the atmosphere (mass of the gas in the atmosphere), which depends on the natural and anthropogenic emissions (Mtons/Year), and the uptake of the gas $(GHGuptake, in\ Mtons/Year)$ which depends on the time constant for each GHG.

For the global carbon cycle, the main equations are the following two equations:

- The net primary productivity in the global biomass, which reflects the carbon uptake from the atmosphere by the biomass:

$$NPP = NPP_0\left(1 + \beta_b \ln\frac{C_a}{C_{a,0}}\right) \times Effect\ of\ Warming\ on\ C_{flux\ to\ biomass}$$

Equation 134

Where NPP_0 is the reference Net Primary Production, β_b is the bio stimulation coefficient and $C_{a,0}$ is the reference C in atmosphere (preindustrial).

- The concentration of carbon in the atmosphere, considering the fluxes of carbon that interact with the atmosphere carbon pool, including the uptake by the biomass and the ocean:

$$\mathrm{C_{conc}}(GtC) = C\ emiss_{CH4oxidation} + FluxC_{biom\to atmosf.} + FluxC_{humus\to atmosf.} + TotalC_{emissions} + FluxC_{permafrost} - FluxC_{atmosf.\to biom.} - FluxC_{atmosf.\to ocean.}$$

Equation 135

Where $\mathrm{C_{conc}}(GtC)$ is the concentration of carbon in the atmosphere, which dependes on 1) the emissions of carbon (total anthropogenic emissions, emissions due to CH_4 oxidation, the flux of carbon from biomass to atmosphere, the flux from hummus to atmosphere, and the carbon emissions from permafrost), and 2) the carbon uptake by sinks (by the biomass and the ocean).

2. The main equations for modelling specific output climate variables are the following:
   - Sea level rise depends on global mean temperature change (main output of the climate module):

$$\text{Sea level rise (SLR)} = \mathrm{f}\,(\Delta T) = \text{Equilib. change}_{\text{sea level}} + \text{Sensitivity}_{\text{SRL to T}} \times \Delta T$$

Equation 136

Where, $\text{Equilib. change}_{\text{sea level}}$ is the equilibrium change in sea level, and $\text{Sensitivity}_{\text{SRL to T}}$ is the sensitivity of sea level rise rate to temperature change.

- Ocean acidification, which also depends on the global mean temperature change:

$$\mathrm{pH} = \mathrm{f}\,(\mathrm{CO_2 conc.}) = \alpha - \beta \times CO2_{concent} + \gamma \times CO2^2_{concent} - \delta \times CO2^3_{concent}$$

Equation 137

Where, α, β, γ, δ are "pH constants" for the function that estimates pH based on CO_2 concentration [3, 4].

Finally, as explained before, the temperature change has been distributed among WILIAM political regions to better quantify climate change impacts:

- Regionalization of climate variables: In this submodule, we included the equation:

$$\Delta Tcreg(°C) = b \times \Delta Tworld(°C)$$

Equation 138

Where used $\Delta Tceg(°C)$ represents the Temperature change by region and climate, b represents the slope values shown in Table 13 of the Deliverable 6.3 (V Pastor et al., 2021) and $\Delta Tworld(°C)$ is the World temperature change.

Therefore, we also performed the following equation:

$$\Delta Treg(°C) = \sum (Climate \times \Delta Tcreg(°C))$$

Equation 139

Where $\Delta Treg$ is the Temperature change by region, computed as the sum of the respective climatic zones percentage ($Climate$) versus the result of Eq. (139). Climate percentages for each Locomotion Region are described in Table 2 of Deliverable 6.3 (V Pastor et al., 2021).

3. Finally, the main equations for the modelling of biophysical climate change impacts are the following:
   - The main equation for modelling the climate change impacts on forest are the following:

$$\Delta NPP(kg/ha \cdot year) = \left[a + b \times \Delta T(°C) + c \times \Delta T^2(°C) + a_1 \times \frac{\Delta CO_2}{\Delta CO_2 + b_1} \right] - NPP_0$$

Equation 140

Where $NPP(kg/ha \cdot año)$ represents the Net Primary Productivity value of each type of forest. The coefficients a, b, c, a_1, b_1 are the adjusted values obtained by regression and calibration procedures for each region and type of forest. T is the regional temperature value in degrees Celsius for each region and ΔCO_2 is the global change in the concentration of carbon dioxide (in ppm).

   - The main equation for modelling the climate change impacts on crop yields is the following:

$$\mathrm{Crop\_yield}\left(\frac{kg}{ha}\right) = Crop\_yield_{reference}\left(\frac{kg}{ha}\right)\left[1 + a_1 \times \Delta T(°C) + a_2 \times (\Delta T)^2(°C) + b_1 \times \frac{\Delta CO_2}{\Delta CO_2 + b_2} \right]$$

Equation 141

Where Crop_yield $\left(\frac{kg}{ha}\right)$ represents the crop yield for different crop types affected by temperature change, while $Crop\_yield_{reference}$ is the reference crop yield. The coefficients a_1, a_2, b_1, b_2 are the adjusted values obtained by regression and calibration procedures. T is the change in temperature, and ΔCO_2 is the global change in the concentration of carbon dioxide (ppm), which refers to the same reference year of the $Crop\_yield_{reference}$.

- The main equation for modelling the climate change impacts on land due to sea level rise based on (Roson & Sartori, 2016) is the following:

$$\mathrm{LRT}_i = \mathrm{LR}_i * SLR_i = \sum_{i=1}^{N} LR_i * (\alpha + \beta \Delta t - V_i)(t - 2000)$$

Equation 142

Where LRT_i is the percent of agricultural land lost for the sea level rise in each country "I" (ha), and LR_i is the effective percent of land change per meter of sea level rise (ha/m)

4.8.3. DATA SOURCES

Most relevant data of the Climate module is described in the following points. Specially the data collected are from 2005 and 2020 intervals:

- Main data for sources used for introducing the emissions and calibrating and validating of GHG cycles and climate variables are the following:
  - Exogenous GHG emissions. Data from RCPs, processed and validated with EDGAR database (European Commission, n.d.). Sources: 1) RCP database and 2) European Commission, Joint Research Centre (JRC)/Netherlands Environmental Assessment Agency (PBL). Emission Database for Global Atmospheric Research (EDGAR)(European Commission, n.d.).
  - Initial and historical GHG concentration, for the start year of the simulation come from the historical data from "Forcings in GISS Climate Model" from NASA (National Aeronautics and Space Administration), Goddard Institute for Space Studies. Available in https://data.giss.nasa.gov/modelforce/ghgases/
  - Global mean CO2 concentration (records). For calibrating and validating CO_2 concentration the data from the Global Monitoring Laboratory from NOAA has been used [19]
  - Global Annual Mean Surface Air Temperature Change. For calibrating and validating global temperature change. Sources:
    - GISS Surface Temperature Analysis (GISTEMP): Analysis Graphs and Plots. Available at: https://data.giss.nasa.gov/gistemp/graphs/# .
    - HadCRUT4 near surface temperature data set. Met Office Hadley Centre observations datasets. Available at: https://www.metoffice.gov.uk/hadobs/hadcrut4/data/current/download.html.
- Main data sources for climate change impacts:
  - Data for calibrating the equations of climate change impacts on crops and on forests were obtained from simulations of the LPJmL model (Schaphoff et al., 2018). These simulation results of the LPJmL model are freely available at ISIMIP database in a geo-located gridded format.

4.8.4. IMPLEMENTED POLICIES

[19] Dr. Pieter Tans, NOAA/GML (gml.noaa.gov/ccgg/trends/) and Dr. Ralph Keeling, Scripps Institution of Oceanography (scrippsco2.ucsd.edu/). Available at https://gml.noaa.gov/ccgg/trends/mlo.html

The Climate module does not contain strictly policies, as it is mainly a module that calculates outputs based on the climate policies applied in the rest of modules (or sectors). However, it includes hypothesis or assumptions that has uncertainty (e.g., have a "likely" range of values) and therefore can be changed. However, the default values are based on literature, so for changing the default values expert knowledge should be applied. The list of hypotheses is the following:

- Climate sensitivity. This parameter takes values in a "likely range". In particular, it reflects the increase of the global temperature with a doubling of atmospheric CO_2 equivalent. The range is based in the last report of IPCC AR6 (Forster et al., 2021), and takes a default value of 3°C, within a likely range between 2°C (high confidence) and 5°C (medium confidence).
- Select RCP for setting the GHG emissions of those gases not being modelled endogenously. There are for options: RCP 2.6, 2. RCP 4.5, 3. RCP, 6.0 4. or RCP 8.5. This allows to set a default RCP (Representative Concentration Pathway) which reflects the pathway of emissions of GHGs in the years to come of those GHG emissions that are not calculated within the scenario of WILIAM. This hypothesis allows to drive the evolution of those dimensions that have not been fully endogenized in WILIAM (e.g., some GHG emissions as for example some specific coming from industry).
- Global Warming Potential (GWP) time frame. This refers to this parameter that allow comparisons of the global warming impacts of different gases relative to the CO_2. As the gases are kept in the atmosphere absorbing energy during a specific time period (depending on the gas) it is necessary to consider a time frame for this comparison and calculation of the forcing of all the GHGs together. There are two options: 20 years or 100 years. By default, 100 year is selected because it is more used, although a shorter time could be also selected (e.g., CH_4 has a much higher short-term warming effect that CO_2).

4.8.5. KEY OUTPUTS

The most relevant outputs of the Climate module are as follows:

- Total GHG emissions (sum and alignment of the emissions coming from the rest of modules)
- GHGs concentrations: "C in Atmosphere", "CH_4 in Atmosphere", "N_2O in Atmosphere", "PFC in Atmosphere", "SF_6 in Atmosphere", and "HFCs in Atmosphere". Units: Gt, or Mtons
- Total Radiative Forcing. Units: watt/(meter*meter)
- Mean global temperature change. Units: DegreesC.
- pH_ocean. Units: pH.
- Global sea Level Rise. Units: mm.
- Temperature change by region and climate; and Temperature change by region, both in degrees Celsius (°C)

4.9. THE MATERIALS MODULE

4.9.1. GENERAL DESCRIPTION OF THE MATERIALS MODULE

The Materials module is structured into (1) the Metal submodules (Fe, Cu, Ni and Al), (2) the Fossil fuels submodules (natural gas, crude oil, and coal) the (3) Uranium submodule, and (4) the submodules of Material requirements of green energy technologies (for solar PV, wind onshore, wind offshore and electric batteries). The Figure 44, Figure 45 and the Figure 46 present the relationship to other modules of WILIAM.

The Metals and Fossil fuel submodules (1-2) work in one global region and mainly interact with the Economy and Energy modules of WILIAM. The main purpose of the Materials module is to model the extraction, and availability of resources and their prices. The price signal is than delivered to the Economy module. In the case of the Fossil fuels submodules, the price signal is part in the calculation of a final energy demand, which is delivered to the Energy module. The Energy module calculates how the final energy demand will be met (different energy technologies) and calculates the primary energy demand by commodity, which is delivered back to the Fossil fuel submodule. In the case of the metals the prices calculated in the Metal submodules, they are linked to the Economy module, where a certain price triggers a certain level of economic activity (calculating final demand), the final demand in monetary units is translated back into material demand (physical units) for different metals.

For uranium (3), an approach based on depletion curves identified in the literature was applied. Depletion curves represent extraction levels compatible with geological constraints as a function of time, which are adapted since uranium demand in WILIAM is endogenous. Hence, there is no feedback with economy for uranium, but just material-energy feedback which constrains nuclear capacity expansion and use when uranium reaches maximum physical extraction. For each of the key "green" technologies, most representative technologies in the market were selected, considering their present and foreseen performance.

The submodules of Material requirements of energy technologies (4) cover two main objectives: first, they are the basis for the computation of the endogenous and dynamic EROI of the green technologies (cf. section, 4.9.5; second, they allow to compare the cumulative primary demand of minerals for green technologies with current estimated mineral endowments (reserves and resources).

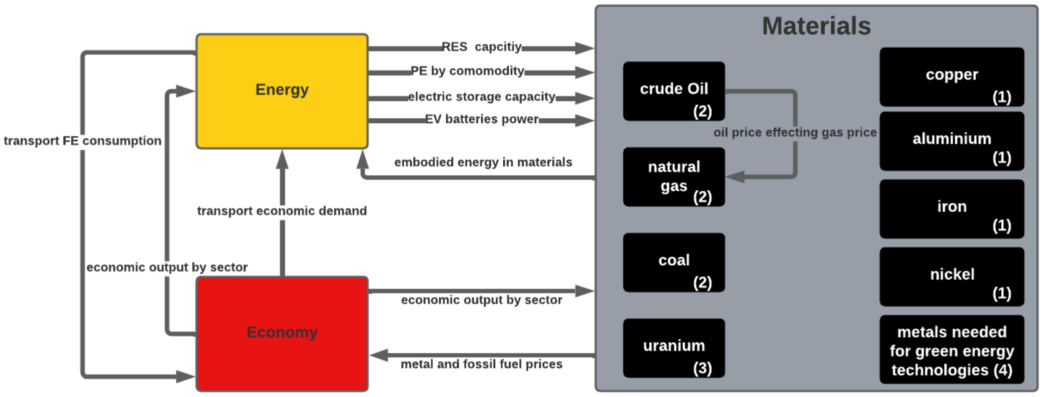


Figure 44: Overview of the Materials Submodules with linkages to other modules of WILIAM.

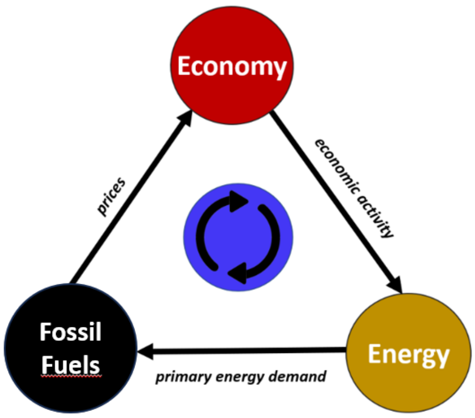


Figure 45: Relation of fossil fuel module with the rest of the WILIAM model

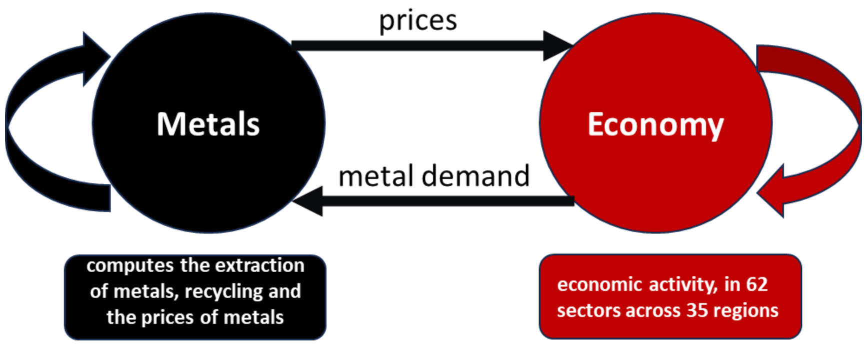


Figure 46: Relation of the Metals Submodules with the rest of WILIAM.

4.9.2. FOSSIL FUEL SUBMODULES

The Fossil fuel submodule in WILIAM consist of three submodules crude oil, natural gas and coal. Their objective is to encapsulate the numerous factors impacting fossil fuel supply: resources, reserves, exploration investments, extraction capacity, actual production, and prices. This module interacts with the Energy and Economy modules, see Figure 45, through a feedback loop based on price and quantity. This cycle initiates in the Economy module, where the demand for 62 goods/services and 35 regions—corresponding to a specific fossil fuel price level—is calculated. This data is relayed to the Energy module, which successively computes final energy demand, operations of diverse energy transformation technologies, which then are resulting in primary energy demand of crude oil, coal, and natural gas. The latter integrates with the Fossil fuel submodule, interacting with supply to determine a new price, production capacity investments, and alterations in resources and reserves. Prices are dictated by the tension between supply and demand: as the disparity between demand and supply grows/shrinks, prices fall/rise. The new price loops back into the Economy module at the next time step, impacting goods and services prices via the supply chains of the multi-regional input-output model. Different households across the 35 regions respond to these price alterations based on a predefined set of income and price elasticities. Trade structures also shift due to the changing price ratio between domestic and imported goods. Ultimately, fossil fuel price fluctuations influence the level and composition of demand for the 62 goods/service types, affecting various aspects of the Economy module—production, investment, wages, employment, income, profits (including from oil extraction), government revenue, and expenditure. These economic alterations subsequently feedback into the Energy and Materials modules. In essence, the link between the energy, materials and economy submodule operates on a demand-price mechanism, where an elevated price 1) decreases fossil fuel demand (due to the economy-energy linkage) and 2) provokes investment in extraction capacity expansion, constrained by available resources and increasing depletion rates. A general overview of the structure of the fossil fuel submodule is displayed in Figure 47.

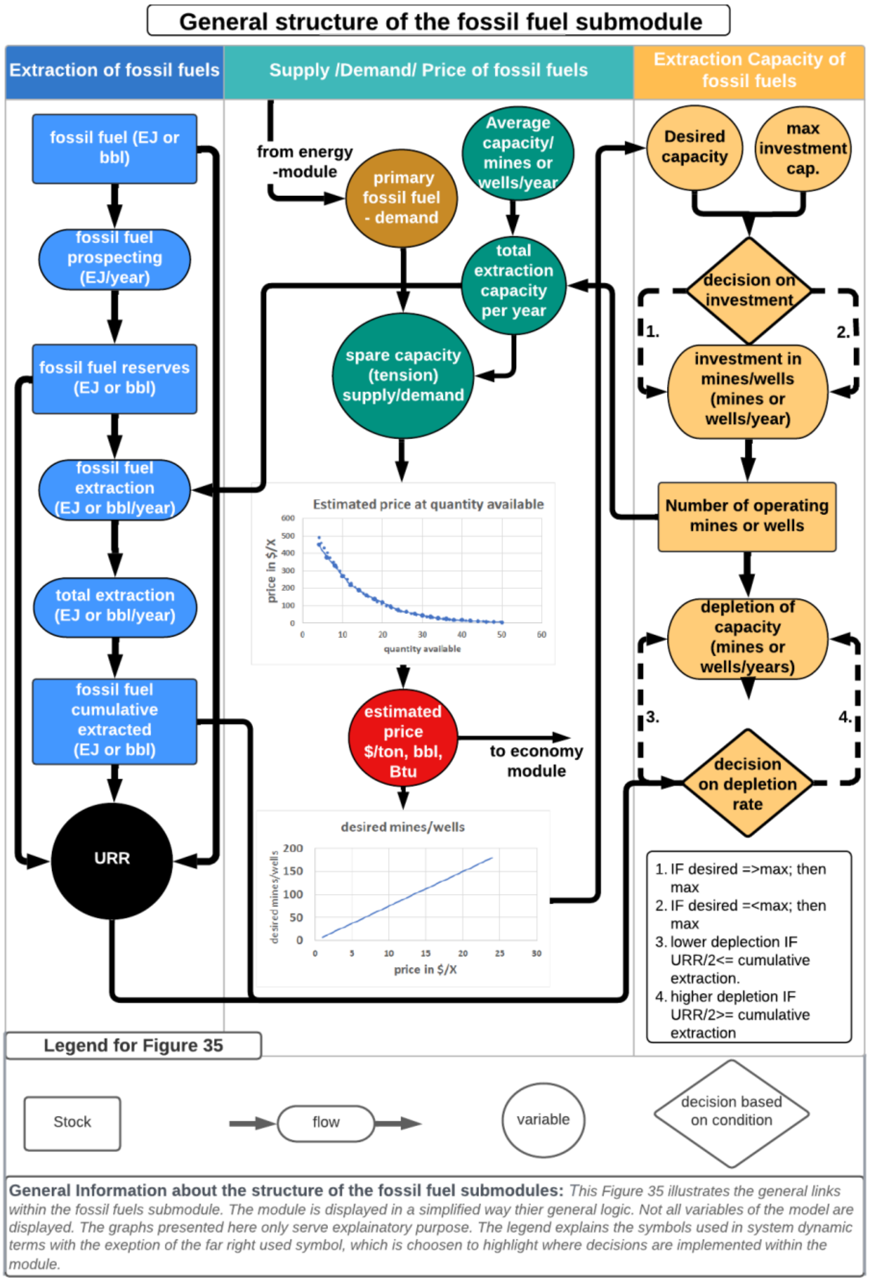


Figure 47: General structure of the fossil fuel submodule.

4.9.2.1. CRUDE OIL SUBMODULE

The oil submodule is developed to compute the oil extraction and oil price the general structure is similar to the structure illustrated in Figure 47. The oil extraction and oil price are modelled on a global level. Oil demand is coming as an input to the oil submodule from the Energy module of WILIAM as a demand for primary Energy in EJ per year. The values of the Energy module get converted into barrels and are used to compute the current oil price depending on the tension between supply and demand. The oil price is sent to the Economy module to compute the final energy demand for the following timestep, this information is passed to the Energy module where in the energy transformation chain a calculation of the final energy demand to the primary energy demand is taking place. The newly calculated primary energy demand is than passed to the again to the oil submodule. The Figure 47 describes in detail the fossil fuel submodules. It is divided into three different sections; extraction of fossil fuels (here oil), which covers the oil resources, oil reserves, cumulatively extracted oil and the ultimately recoverable resources. The second section covers the supply, demand, fossil fuel price (here oil). The third section covers the investment in fossil fuel (here oil) extraction infrastructure, in the case of oil the investment in extraction infrastructure takes place in form of investment in oil wells, which is compared to the historical well count for the time frame of 2005 to 2020, and also accounts for wells that are getting depleted.

4.9.2.2. NATURAL GAS SUBMODULE

The gas submodule has been designed to calculate gas extraction and determine the gas price on a global scale (as depicted in the general overview of the fossil fuel module, see Figure 47). Demand for natural gas is introduced to the natural gas module from WILIAM's Energy module, inform of primary Energy in Exajoules (EJ) per year. The natural gas price is calculated in $ per million Btu, based on the tension between supply and demand. The calculated gas price is relayed to the economic module, which computes the final energy demand for the subsequent timestep. This data is then routed back to the Energy module where, within the energy transformation processes, the final energy demand is retranslated into the primary energy demand. This recalculated primary energy demand is subsequently reintroduced to the gas submodule. The comprehensive structure of the gas submodule is illustrated in the Figure 47, which illustrates the general structure of the fossil fuel module. It is divided into three distinct sections. The first section deals with the extraction of gas, covering aspects such as gas resources, gas reserves, cumulative gas extraction, and the ultimately recoverable resources. The second section delves into the supply and demand dynamics of gas, the gas price, and investment geared towards gas extraction infrastructure. Finally, the third section pertains to the gas extraction infrastructure, including the operation of active extraction capacity and the process of extraction capacity depletion. Here the extraction capacity got matched with the historical extraction, since there was not the same quality of data available regarding active producing wells like in the oil module.

4.9.2.3. COAL SUBMODULE

The Coal submodule has been designed to calculate coal extraction processes and establish the coal price at a global level (see general structure of fossil fuel module, Figure 47). The demand for primary Energy, expressed in Exajoules (EJ) per year, is supplied to the coal model by WILIAM's Energy module. These energy values are recalculated into tonnes and serve in computing the current coal price per ton, governed by the interplay between supply and demand forces. This calculated coal price is then transferred to the Economy module to calculate the subsequent timestep's final energy demand. This information is rerouted to the energy module, wherein the energy transformation chain translates the final energy demand back into primary energy demand. The re-evaluated primary energy demand is then fed back into the coal submodule. The structure of the coal submodule is outlined in the Figure 47. The coal module is sectioned into three parts; the first part entails coal extraction, covering coal resources, coal reserves, and cumulatively extracted coal. Coal resources, reserves and active mines are separated into hard and brown coal following the classification of Rogner et al., 2012 , these coal resource and reserve stocks are separated into the categories, hard and brown coal. This section also incorporates the ultimately recoverable resources. The second part revolves around the supply and demand dynamics of coal, the pricing of coal, and investments made towards coal extraction infrastructure (mines). The mines and investment towards the mines are also divided into the two categories hard and brown coal. The final part in yellow focuses on the physical infrastructure for coal extraction, comprising active mines (separated into hard and brown coal mines) and those nearing depletion.

4.9.3. METAL SUBMODULES LINKED TO ECONOMY

The Metal module consist of the submodules of (Fe, Al, Cu and Ni) and is a System Dynamics module based on mass balance expressed through differential equations. Metal mining's basic driving mechanism originates from profitability and the availability of a mineable reserve, as used in the model. This factor is influenced by mining costs, capital costs, and ore grade.

The mining process of Cu, Al, Fe and Ni differs, some of the metals are just mined from primary sources like iron and aluminium while others such as nickel and copper are also mined from secondary metals as a co-product, adding another dimension to the extraction process.

In the module, some metals are divided into up to six ore qualities: rich quality, high-quality, low quality, ultralow quality, trace quality and ocean ore grade quality, while other metal like Al are only divided into two ore grade qualities.

The higher the quality the higher the metal content in the host material. A higher quality is also associated with lower cost for extraction and production of metal. A lower ore grade implies that more rock must be moved to mine the metal, meaning that a higher metal price is needed to maintain production levels. The price must exceed production costs and is determined by supply and demand tension. Supply encompasses the amount of metal moving to the market plus the amount of metal in the market stock, with demand coming from the Economy module. The quantity in the trade market depends on the balance between metal production deliveries and the amounts recycled from metal scrap. This profit is driven by the metal price and the amount extracted, offset by extraction costs. The price in the module is based on econometric estimations and is calibrated to follow the trend of the historical price. For metal extraction to continue, profit must exceed production costs.

There are two possible setting for the recycling rate, in one setting, the price also drives the urge for metal recycling (price affecting the recycling rate), in the second setting the user has the option to set an exogenous recycling rate. The price of metal is linked to the Economy module, corresponding to the sectors 'IRON/ ALUMINIUM/ COPPER or NICKEL MINING. Demand comes from historical data for the timeframe from 2005 until 2015 and from the Economy module from 2015 onwards (sum of ''IRON/ ALUMINIUM/ COPPER or NICKEL MINING recalculated from monetary to physical units).

Based on the demand, the Metal module calculates the available metal and its price. The new price then feeds back into the Economy module at the next time step, impacting goods and services prices via the supply chains of the multi-regional input-output model. Different households across the 35 regions respond to these price changes based on a predefined set of income and price elasticities. Trade structures also shift due to the changing price ratio between domestic and imported goods. Ultimately, metal price fluctuations influence the level and composition of demand for the 62 goods/service types, affecting various aspects of the Economy module—production, investment, wages, employment, income, profits (including from metal mining), government revenue, and expenditure. These economic changes subsequently lead to the calculation of new metal demand, which feeds back into the corresponding Metal submodules.

4.9.3.1. COPPER SUBMODULE

The Figure 48, presents a simplified flow chart for copper. The Copper submodule is a System Dynamics model based on mass balance expressed through differential equations. The most relevant stocks and flows of the Copper submodule are as follows:

1. Extractable amounts
    a. Rich, high, low, ultra-low, trace grade and oceans grade quality ore
        i. Resources
        ii. Reserves
2. In society

a. copper market stock
b. stock-in-use copper in society
c. scrapped copper not yet lost or recycled.

The basic driving mechanism of copper mining comes from profits and availability of a mineable reserve used in the module, but affected by the mining cost and how that is modified with capital costs, and ore grade. In the copper submodule, copper ore is divided into six ore qualities: rich quality, a high-quality grade. Low quality, ultralow quality, trace quality and ocean ore grade. The higher the quality the higher the Cu content in the host material. A higher quality is also associated with lower cost for extraction and production of copper. Another speciality of the copper submodule is that copper is also mined from secondary metals as a co-product, adding another dimension to the extraction process. In the copper submodule has two options for the recycling function, .in one module setting, the price also drives the urge for metal recycling (price affecting the recycling rate), in the second option the user also has the option to set an exogenous recycling rate. The copper model has only one stock for copper in use in society, unlike Iron where the use is divided into different categories like short, medium, and long-term uses. Despite these differences the general description from section 4.9.3 applies for the copper module.

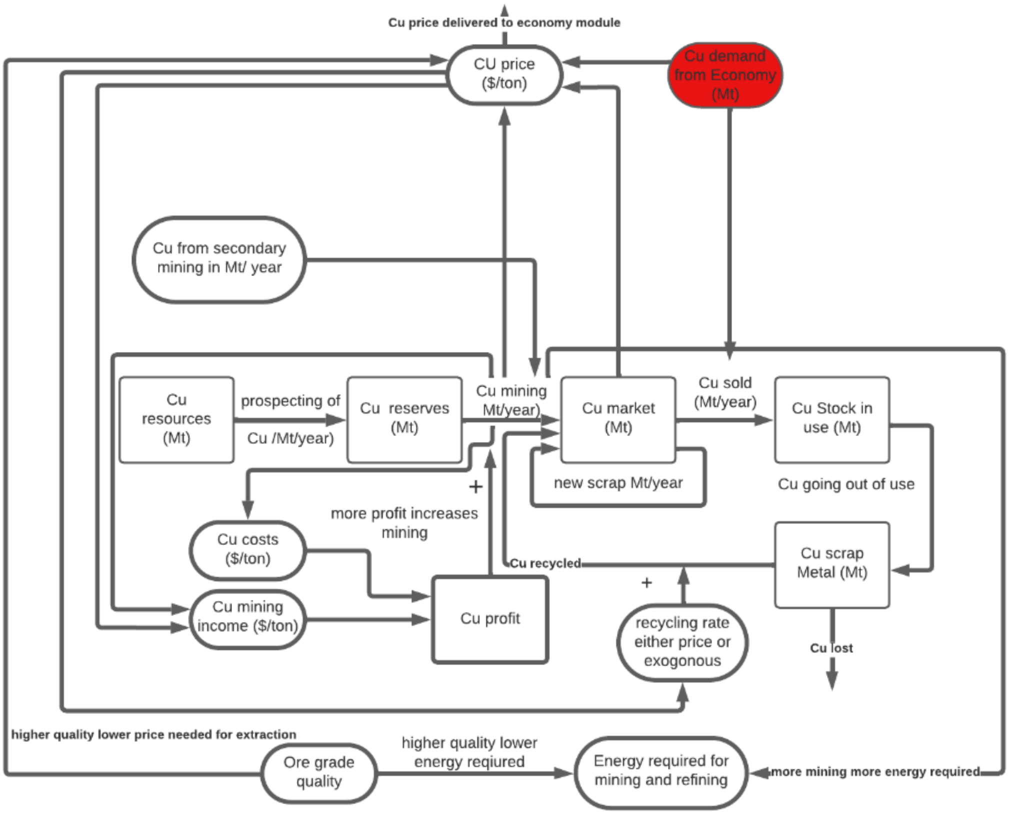


Figure 48: Copper submodule.

4.9.3.2. ALUMINIUM SUBMODULE

The Figure 49 shows a simplified flow chart for aluminium. The Aluminium submodule is a System Dynamics model based on mass balance expressed differential equations. The most relevant stocks and of the aluminium submodule:

3. Extractable amounts aluminium
    a. High quality ore expressed as aluminium (low silica content)

i. resources
ii. reserves

b. Low quality ore expressed as aluminium (higher silica content, some contaminants)
   i. resources
   ii. reserves

4. In society
   a. aluminium market stock
   b. stock-in-use aluminium in society
   c. scrapped aluminium not yet lost or recycled.

The basic driving mechanism of aluminium mining comes from profits and availability of a mineable reserve used in the module, but affected by the mining cost and how that is modified with capital costs, and ore grade. In the model, the aluminium ore is divided into two qualities: one high quality bauxite grade with low content of silica as an impurity in the alumina; and the other one low quality with higher silica content, making it more expensive to reduce to aluminium. It consists of low-grade bauxite and other types of low-grade alumina and nepheline.

The price can in one setting of the model also drives the urge for recycling of aluminium (price effecting the recycling rate), but the user also the possibility to set an exogenous recycling rate. The price of aluminium is linked to the Economy module, matching the sector aluminium MINING AND MANUFACTURING ALUMINIUM. The calculation of the demand and the feedback follows the same principal as descripted in the general section for the Metal submodules cf. 4.9.3.

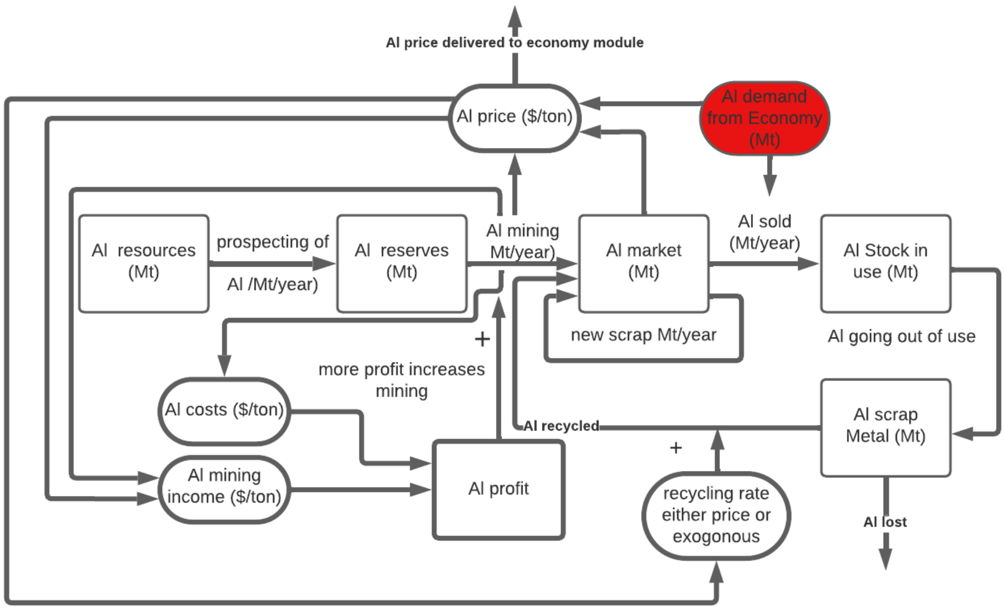


Figure 49: Aluminium submodule.

4.9.3.3. IRON SUBMODULE

The Figure 50 shows a simplified flow chart for **iron**. The **Iron** submodule is a System Dynamics module based on mass balance expressed differential equations. The most relevant stock and of the **Iron** submodule:

1. Iron
   a. Mineable stocks.

i. Known; rich-, high-, low- and ultralow grade ore.
ii. Hidden; rich-, high-, low- and ultralow grade ore.

b. In society, we distinguish.
   i. Fe Trade market
   ii. Stock-in-use
      1. Long term
      2. Intermediate term
      3. Short term
      4. In Stainless steel
   iii. Scrapped

The Iron module possesses distinct characteristics. Notably, its resource and reserve stocks are categorized into four ore grade classifications: rich, high, low, and ultralow grade ore. Additionally, unlike other metal modules, the iron stock in use is delineated into short, medium, and long-term utilization. Beyond these distinctions, the module's functionalities and feedback mechanisms align closely with those of the Al, Cu, and Ni modules. A comprehensive description is available in section 4.9.3.

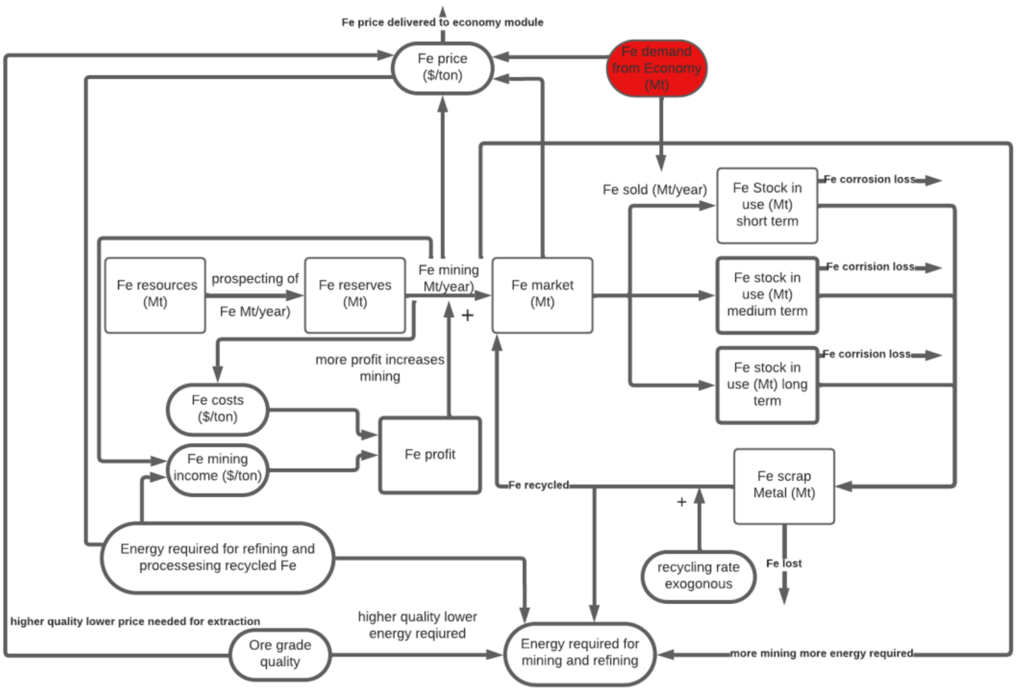


Figure 50: Structure of the Iron submodule.

4.9.3.4. NICKEL SUBMODULE

The Figure 42 illustrates a streamlined flow diagram concerning nickel. Within this context, the Nickel submodule operates as a System Dynamics model, applying mass balance principles expressed via differential equations. This submodule encompasses various key stocks and flows, detailed as follows:

1. Nickel
   a. Mineable stocks.
      i. Ni resources separated into rich-, high-, low- and ultralow and trace grade ore.
      ii. Ni reserves separated into rich-, high-, low- and ultralow and trace grade ore.
   b. In society, we distinguish.
      i. Trade market
      ii. Stock-in-use
         1. as plating
         2. other

3. In stainless steel

iii. Scrapped

The foundational mechanism driving nickel mining is rooted in both profitability and the presence of reserves that can be mined. This driving force is modulated by factors including mining expenditures, capital outlays, and the grade of the ore. Unlike metals like iron and aluminum, nickel's extraction is often coupled with the mining of secondary metals, forming a co-product, and introducing an additional facet to the mining process. In the given submodule, nickel ore is classified into six distinct qualities: rich quality, high-quality grade, low quality, ultralow quality, trace quality, and ocean ore grade. The concentration of nickel (Ni) in the host material increases with the quality of the ore, and higher quality ore typically leads to reduced extraction and production costs.

Distinct from the Metals modules of Fe, Al, and Cu, the stock of Ni in use is categorized into three segments: Ni in other use, Ni plating, and Ni for stainless steel. Each segment has varying proportions of the sold Ni channeled into it. Notably, Ni plating and Ni in other use have different scrapping rates. Another unique feature of this module, in contrast to other metal submodules, is that its recycling rate is either exogenously fixed or determined by the user. Despite these variations, the module's operations and feedback loops are consistent with the Al, Cu, and Fe modules. See more details in section 4.9.3.

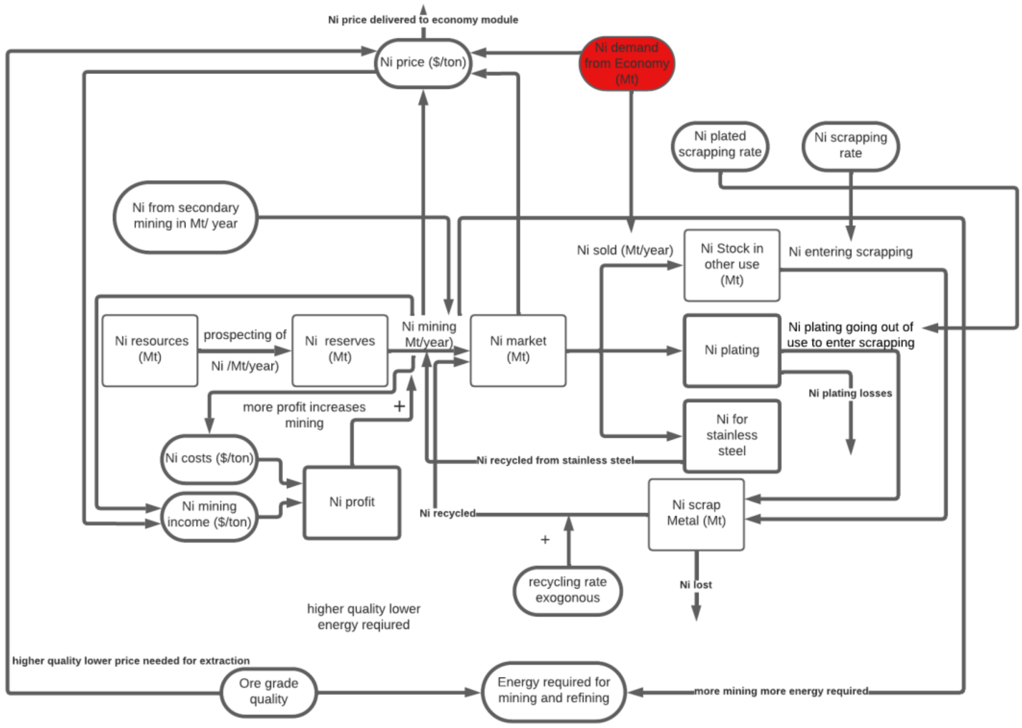


Figure 51: Nickel submodule.

4.9.4. URANIUM SUBMODULE

Uranium submodule communicates with the energy module directly without direct interaction with the economy module. In this case, a stylized approach based on depletion curves identified in the literature is applied. These depletion curves represent extraction levels compatible with geological constraints (both stock and flow) as a function of time, which are adapted since uranium demand in WILIAM is endogenous. Depletion curves, rather than predictions, represent maximum extraction profiles for a fraction of the resource base estimated to be economically extractable in the future considering geological constraints: the actual rate of resource consumption might be affected by such variables as geopolitics, economic conditions and technology development. Given that demand is endogenously modeled for each resource, these depletion curves are transformed so as to be incorporated in the dynamic model, being converted into maximum production curves as a function of remaining resources. We assume that, while the maximum extraction rate (as given by the depletion curve) is not reached, the extraction of each resource matches the demand. Actual extraction will therefore be the minimum between the demand and the maximum extraction rate. To do this, the depletion curves have been converted into maximum production curves as a function of remaining resources. In these curves, as long as the remaining resources are large, extraction is only constrained by the maximum extraction level. However, with cumulated extraction, there is a level of remaining resources when physical limits start to appear, and maximum extraction rates are gradually reduced. In this way, the model uses a stock of resources (the RURR) and it studies how this stock is exhausted depending on production, which is in turn determined by demand and maximum extraction. The model user is free to select any depletion curve already loaded in the model, introduce a new one, or even assume that no relevant constraints will affect the supply of uranium in the simulation period (Figure 52).

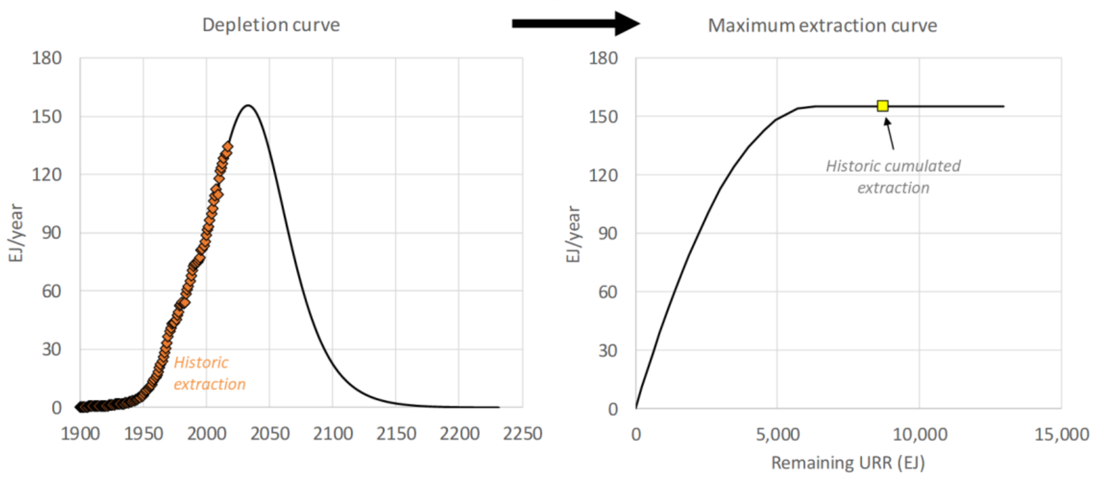


Figure 52: Example of the transformation of a depletion curve (left) in maximum extraction curve (right) suppressing time dependence as a function of remaining ultimately recoverable resources (URR). Source: (Capellán-Pérez et al., 2020).

4.9.5. MATERIAL REQUIREMENTS OF GREEN ENERGY TECHNOLOGIES

Given the large amount of energy technologies, focus is on the main green energy technologies for electricity generation (solar and wind) and electric batteries as having been identified in the literature as dependent on potentially critical minerals in the future. For solar CSP, wind onshore and wind offshore, a representative technology was selected, considering the present and foreseen performance while avoiding those more likely to be affected by scarce minerals in the future. For solar PV and electric batteries, a more in-depth analysis was conducted directed to identify the most relevant sub technologies given their very different material requirements: mono-Si, poly-Si, CIGS and CdTe for solar PV (including differentiating between rooftop and on ground installations), and LMO, NMC622, NMC811, NCA and LFP for electric batteries, respectively. For electric mobility, the chargers, additional grids and materials for railways catenaries were also covered. For stationary batteries, we assume the LFP technology.

Subsequently, a literature review was performed to identify the material intensity (kg/MW) of each technology, including those related to additional grid requirements, for a total of 24 minerals: aluminum, cadmium, chromium, cobalt, copper, gallium, graphite, indium, iron, lithium, magnesium, manganese, molybdenum, neodymium, nickel, lead, selenium, silver, tin, iron/steel, tellurium, titanium, vanadium, and zinc. Selection criteria were based on considering all relevant materials so as to accurately estimate the embodied energy for the EROI estimation, with potential critical minerals having already been identified in the literature, as well as specific assessments. A comprehensive literature review, complemented by our own estimations, was performed to collate the most robust and accurate data concerning material requirements for each technology. In the case of uncertainty about potential double accounting, material requirements were not included. Hence, these estimations can be considered conservative/optimistic.

For these 22 metals the primary demand of minerals (after accounting for recycling rates) is compared with the current estimated level of their geological availability (reserves and resources). Hence, potential mineral scarcity in the current model version is not feed-backed and does not affect the rest of the model, as is the case with the metal submodules linked to economy (cf. section 4.9.3 copper, aluminum, iron and nickel). Note that there is some overlapping for these 4 base metals, in the model these submodules work independently, and the bottom-up material requirement of energy technologies is used for a qualitative detection of risks to the future mineral supply. In theory, the approach used for the 4 based metals could be extended to other minerals, but this is hindered by the lack of specific sectoral disaggregation in the economy module (cf. section 4.4.3.3 the Production/Firms submodule). For the minerals not covered by the metal submodules linked to economy, a stylized approach is included in order to compute the demand of minerals for the rest of the economy, assuming a linear dependence with GDP. Historical data for the global GDP and the extraction of minerals for the period 1994-2015 were applied to estimate the parameters of the regressions, obtaining acceptable correlations in most cases.

4.9.6. METHODOLOGICAL APPROACH

For each of the resources, systems analysis was employed to conceptually model the production, market, stocks in use and consumption system. The flow pathways, causal chains, and the entire feedback structure of the system were mapped using causal loop diagrams and flow charts. The resulting coupled differential equations were converted into computer codes for numerical solutions.

Initially, for all the resources, a stand-alone module was crafted and tested. Once that model operated satisfactorily, it was integrated as a submodule in the WILIAM model structure. The model was first utilized to reconstruct the past (2005-2020) to assess its performance and robustness. When the results were satisfactory, the model was then used to simulate the potential future until 2050 or 2100, grounded in its capability to reconstruct the observed past pattern for 15 years.

The model was developed iteratively, continually testing it against data sourced from academic literature and corporate databases. This methodology was consistently applied throughout the study and during the model's development. The iterations were crucial in setting parameter values that could reproduce the mining or extraction history, observed ore grades, and price structures. This approach enabled us to identify the system's intervention points and suggest policy interventions.

All the material modules, Fossil fuel and Metals are modelled on a global scale. In the following sections we will list the most relevant equations of the Metal and fossil fuel module.

4.9.6.1. EQUATIONS FOR THE METAL MODULE

The metal modules are based on coupled differential mass balance equations. In addition to the mass balances comes equations for ore grades, recycling, profit, and price use. The ore is mined from the grade with the lowest extraction cost. Dependent stocks are the resource, the reserve, the metal market, the metal in use in society and the metal in scrap stock. The mining activity is profit driven, the profit is affected by the mining cost and the market price. A lower ore grade implies that more rock must be moved to mine the metal. All models operate on a global scale and the mining and recycling activity captures the global amounts mined and recycled:

Iron, Copper, Aluminium and Nickel module

The Iron, Copper, Aluminium and Nickel modules share similar equations for mining and prospecting.

Mining Equation:

$$r_{Mining} = k_{\ change\ in\ grades} * m_{reserves} * f(Technology) * g(profit) * p(yield)$$

Equation 143

where r_{Mining} is the rate of mining, k is the rate coefficient depending on the current price ramping up extraction of lower ore grade with rising price and m is the mass of the ore body (reserve size of that ore grade), where f(Technology) is a technology improvement function dependent on time, g(profit) is a coefficient influenced by feed-back from profits and is affecting the mining rate. The p(Yield) is a rate adjustment factor to account for differences in extraction yield when the ore grade decreases; *this factor is not present in the aluminium model*. These functions are given exogenously to the model. The rate coefficient is modified with ore extraction cost and ore grade. The size of the extractable ore body (reserves), $m_{reserves}$, is determined by the extractions (r_{Mining}) and prospecting ($r_{Discovery}$).

Reserve stock equation:

$$\frac{dm_{reserves}}{dt} = -\,r_{mining} + r_{discovery}$$

Equation 144

The amount of reserves depends on discovery and the amount mined. The amount of resources ($m_{resources}$) decrease with the rate of mining (r_{mining}) and the rate of discovery which depends on the amount of metal resources ($m_{resources}$) and the prospecting coefficient $k_{Prospecting}$. The prospecting coefficient depend on the amount of effort spent and the technical method used for prospecting.

Prospecting equation:

$$\frac{dm_{resources}}{dt} = -r_{Discovery} = -k_{Prospecting} * m_{resources} * f(Technology)$$

Equation 145

The amount of resource ($m_{resources}$) decreases with the rate of discovery ($r_{Discovery}$). The rate of discovery is dependent on the amount of metal resources available ($m_{resources}$) and the prospecting coefficient $k_{Prospecting}$. The prospecting coefficient depends on the amount of effort spent and the technical method used for prospecting. Additionally, the prospecting is improving over time based on the parameter, f(Technology) which is based on a technology improvement function dependent on time.

Prospecting is done to find unknown but existing resources, to convert them to known and extractable reserves. When the prospecting efficiency decrease because there is less resources left to find, the prospecting will be broken off. The prospecting rate is set as an external parameter to match historical data.

Corrosion and Losses in the Iron and Nickel Module:

The rate of corrosion of the stock-in-use ad from scrapped metal are defined as:

$$r_{Corrosion} = -\,k_{Corrosion} * m_{Stock-in-use}$$

Equation 146

Scrap is both lost physically by dropping it where it cannot be found and by corrosion:

$$r_{Scrap\ loss} = -(k_{Scrap\ loss} + k_{Corrosion}) * m_{scrap\ stock}$$

Equation 147

Iron and Nickel from mining to market:

The iron ore beneficiation and subsequent smelting yield is defined as:

$$r_{smelting\ supply} = k_{Smelting\ yield} * r_{Mining}$$

Equation 148

For nickel the equation is somewhat different:

$$r_{Nickel} = r_{Nickel\ mining} + r_{Ni\ from\ PGM} + r_{Ni\ from\ Cu}$$

Equation 149

For nickel, the equation includes secondary sources from copper mining and platinum group metal mining.

Recycling and Metal going out of use in the Aluminium and Copper module:

$$r_{going\ out\ of\ use} = r_{in\ use\ in\ society} * r_{scrapping\ rate}$$

Equation 150

The amount of metal leaving the metal in use stock is a function based on the amount of metal in use times the specific scrapping rate, $r_{scrapping\ rate}$, which differs for aluminium and copper and is given exogenously.

Metal in scrapping for aluminium and copper

$$m_{\ scrap\ stock} = r_{going\ out\ of\ use} - r_{recycled} - r_{metal\ losses} - r_{losses\ during\ prosssing}$$

Equation 151

The $m_{scrap\ stock}$ is the stock of metal that left the in use in society stock and has in the case of aluminium and copper three outflows and one inflow. Metal is entering from the metal in use stock over the flow $r_{going\ out\ of\ use}$. The $m_{scrap\ stock}$ is reduced by metal leaving the stock by metal getting recycled ($r_{recyceled}$), by metal getting lost and not properly recycled ($r_{metal\ losses}$) and by metal that is lost during processing of scrap metal ($r_{losses\ during\ prossing}$).

Metal recycled and sent to the market for aluminium, iron, nickel and copper

$$r_{recycled} = m_{scrap\ stock} * r_{recycling\ rate}$$

Equation 152

The metal flow that gets recycled is determined by the amount of metal available in the $m_{scrap\ stock}$ times the recycling rate, $r_{recycling\ rate}$. The recycling rate is either determined by a function that is dependent on the price of the metal, with a higher price we observe increased recycling rate. In another setting the recycling rate can also be given exogenously by the user. In the nickel module the recycling rate is either a fixed exogenously set parameter or a parameter defined by the model user.

Metal losses during separation and processing for aluminium and copper

$$r_{losses\ during\ processing} = r_{seperation\ loss\ rate} * m_{scrap\ stock}$$

Equation 153

The flow of losses during processing is defined by the rate of loss during the separation process, $r_{seperation\ loss\ rate}$, multiplied with the metal available in the scrap stock, $m_{scrap\ stock}$. The separation loss rate is given exogenously and is based on literature reviews, studying metals flow and recycling efficiency.

Metal lost or not correctly disposed for aluminium and copper.

$$r_{metal\ losses} = m_{scrap\ stock} * (1 - r_{recycling\ rate} + r_{separation\ loss\ rate})$$

Equation 154

The flow of metal lost and not recycled is defined by the amount of metal available for recycling in the $m_{scrap\ stock}$ multiplied with 1 subtracting the current recycling and separation loss rate. This Equation captures metal that never enter proper recycling.

Metal profit (Al, Cu, Ni and Fe)

$$Profit = Income\ from\ sales - mining\ costs$$

Equation 155

The profit for all four modules is defined as the income from sales, which is the mined amount multiplied by the current price subtracting the mining costs, which is essentially the cost of mining and refining a certain ore grade multiplied with the amount mined from that ore grade.

Metal price (Al, Cu, Ni and Fe)

$$estimated\ Price = +EXP(a + b * LN(1/(1 - Q/Q^*)))$$

Equation 156

The calculation listed above is used to calculate the price for the respective resource. The parameters a and b are econometrically estimated, the parameters are chosen to best fit historical price data between (2005 until 2020) of the respective resource. The parameter Q is the current consumption or demand, the parameter Q* is the current extraction capacity.

4.9.6.2. MOST RELEVANT EQUATIONS FOR THE FOSSIL FUEL MODULE

The following equations are the most relevant equations for the three fossil fuels modules and are similar across the crude oil, natural gas and coal model.

Fossil Fuel estimated price calculation:

$$estimated\ Price = +EXP(a + b * LN(1/(1 - Q/Q^*)))$$

Equation 157

The calculation listed above is used to calculate the price for the respective resource. The parameters a and b are econometrically estimated, the parameters are chosen to best fit historical price data between (2005 until 2020) of the respective resource. The parameter Q is the current consumption or demand, the parameter Q* is the current extraction capacity.

Desired number of wells, mines or extraction capacity equation

$$desired\ number\ of\ wells, mines\ or\ extraction\ capacity = (a + b * estimated\ Price)$$

Equation 158

The desired number of new wells/mines/ extraction capacity is calculated as a linear function of the estimated price in the Supply/Demand price block of the model, see Figure 35. The assumption of the authors is that each mine or well has the same productivity, which in reality is not true. This is a simplification of the real world. To econometrically estimate the parameters a and b of the linear function, the historical price and the historical development of the number of active wells and mines were used. This practice was applied for coal and crude oil where mine and well data were available. In the case of natural gas extraction, extraction capacity was used, and the best fit with the historical production and price determined the choice of parameters a and b.

Current extraction of crude oil or coal

$$current\ extraction\ capacity = (nº\ of\ wells\ or\ mines * average\ production\ per\ well\ or\ mines * (1 - actual\ spare\ capacity)$$

Equation 159

The calculation of the current extraction of coal or crude oil is based on the number of active mines or wells times the average production per mine or well, times 1 – the actual spare capacity. The actual spare capacity is determined by the following function:

$$actual\ spare\ capacity = 1 - fossil\ fuel\ demand / fossil\ fuel\ max\ extraction\ capacity$$

Equation 160

The fossil fuel demand is coming from the energy transformation module and the fossil fuel max extraction capacity is determined by the number of active wells or mines and their average productivity per mine.

Current extraction of natural gas

$$current\ Gas\ \ capacity\ = (Gas\ max\ extraction\ capacity * (1 - actual\ spare\ capacity\ gas))$$

Equation 161

In the case of gas, the current extraction calculation is a bit simpler, since there was no reliable well data available. The current gas extraction is determined by the Gas maximum extraction capacity times one minus the actual spare capacity.

4.9.7. DATA SOURCES

4.9.7.1. DATA SOURCES FOR THE FOSSIL FUEL MODULE

In the subsequent sections for the crude oil, natural gas, and coal models, we list the key data used for model calibration and validation. Additionally, we detail the reserve and resource assumptions for various scenarios, along with their sources.

- **The crude Oil Model**

Table 2 include the data sources used for calibration and validation for the timeframe 2005-2020.

Table 2: Data sources for the Oil module.

Historical oil demand	bp-stats-review-2021
Historical oil price	U.S. Energy Information Administration, 2022
Historical wells active	https://www.rystadenergy.com/energy-themes/supply-chain/wells/well-cube/ accessed: 28.07.2022
Historical wells added	https://www.rystadenergy.com/newsevents/news/newsletters/OfsArchive/ofs-november-2018/ accessed: 28.07.2022

	https://www.drillingcontractor.org/670000-wells-to-be-drilled-through-2020-28709/ accessed: 28.07.2022 https://www.rystadenergy.com/energy-themes/supply-chain/wells/well-cube accessed: 28.07.2022
OPEC spare capacity	https://www.eia.gov/finance/markets/crudeoil/supply-opec.php accessed 28.07.2022

Scenario parameters for the remaining resources and reserves of materials (RURR) of crude oil are included in Table 3.

Table 3: Scenario parameter for RURR of crude oil.

Units in Gb	Low URR	Medium URR	High URR	Source
Cumulative extracted	1169.9	1169.9	1169.9	(Laherrère, Hall, & Bentley, 2022) recalculated to 2005
Resource	61.5	1812.6	5313.5	Selected range by Authors
Reserves	2270.0	2270.0	2270.0	BGR, 2020 recalculated to 2005
Total URR	3501.4	5252.5	8753.4	

- **The COAL Model**

Table 4 present the data sources used for calibration and validation for the timeframe 2005-2020.

Table 4: Data sources for the coal module.

Historical coal demand	IEA (2020)
Historical coal price	Average Coal price of for each year based on the Source: accessed 10.10.2022 - https://tradingeconomics.com/commodity/coal
Historical active mines	Estimated number of operating mines, by calibration to historical data. Calculation is based on the number of mines in the year 2022 and the amount of mined coal. Then calculated backwards to match the historical data. From the source: "Global Coal Mine Tracker, Global Energy Monitor, July 2022 release.

Scenario parameters for the remaining resources and reserves of materials of coal are included in Table 5.

Table 5: Scenario parameter for RURR of coal.

Units in EJ	Low URR	Medium URR	High URR	Source
Cumulative extracted	2656.88	2656.88	2656.88	(Rogner, 2012)
Resource hard coal	273736.4	391052	508367.6	((Rogner, 2012) -medium Scenario Low scenario: 70% medium scenario High scenario: 130% medium scenario
Resources brown coal	31269.7	44671	58072.3	
Reserves hard coal	18246	18246	18246	
Reserves brown coal	2.78E+03	2.78E+03	2.78E+03	
Total URR	328689	459405.9	590122.8	

- **The natural GAs Model**

Table 6 describes the data sources used for calibration and validation for the timeframe 2005-2020.

Table 6: Data sources for the coal module.

Historical gas demand	BP statistic report 2022
Historical gas price	Historical Gas price average over the different price region. Taken from BP Stat. Review. 2022.

Scenario parameters for the remaining resources and reserves of materials of natural gas are included in Table 7.

Table 7: Scenario parameter for RURR for the natural gas module.

Units in EJ	Low URR	Medium URR	High URR	Source
Cumulative extracted	2920.38	2920.38	2920.38	((BGR), 2017); (Wang & Bentley, 2020)
Resource	15000	23800	50000	((BGR), 2017); (Wang & Bentley, 2020) medium Scenario Low scenario: 70% medium scenario High scenario: 130% medium scenario
Reserves	8678.58	8678.58	8678.58	
Total URR	26598.96	35398.96	61598.96	

4.9.7.2. DATA SOURCES FOR THE METAL OF MODULES

The Metal submodules of Iron, Aluminium, Nickel, and Copper are adapted modules, originating from already published modules as part of the WOLRD 7 model. All the modules got adapted, simplified modified to be compatible for the WILIAM model. The counterparts of the modules in the WOLRD 7 model run over a different timeframe than the WILIAM model, the INITIAL reserve, and resources estimations in the year 2005 that are used for the Metal Submodules stem from the simulations carried out in the WORLD 7 modules. A more detailed description of the methodology and data of each of these modules can be found in the following publications.

<u>Aluminium</u>

- Data from (Sverdrup, 2014a) and (Sverdrup H. U., 2015).

<u>Copper</u>

- Data from **Fuente especificada no válida.**, (Sverdrup H. &., 2014a) and **Fuente especificada no válida.**.

<u>Nickel</u>

- Data from (Olafsdottir, 2021) and (Sverdrup H. U., 2019).

<u>Iron</u>

- Data from (Sverdrup H. U., 2017) and (Sverdrup H. U., 2019).

4.9.7.3. DATA SOURCES FOR THE URANIUM AND MATERIAL REQUIREMENTS OF GREEN ENERGY TECHNOLOGIES MODULES

- Material requirements of main green energy technologies for electricity generation: (de Castro & Capellán-Pérez, 2020; Pulido-Sánchez, 2022) for solar PV sub-technologies.
- Material requirements of electric batteries for transport, including auxiliary elements such as chargers and grids, also railroads catenaries: (Pulido-Sánchez et al., 2022).
- Uranium: depletion curves from (Energy Watch Group, 2006, 2013; Zittel, 2012).

4.9.8. IMPLEMENTED POLICIES

The Materials module contains several policies that affect the main outputs of the module, and they are listed as follows:

- Change the oil resources estimations to low, medium, high and user defined.
- Change the coal resources estimations to low, medium, high and user defined.
- Change the gas resource estimations to low, medium, high and user defined.
- Set a OPEC target price low, medium, high and user defined.
- Set OPEC target quantities that are released or taken from the oil market.
- Tax rate on extraction of resources.
- Reduction rate material intensity PV panels.
- Set EOL recycling rate for the modelled metals.
- Select Uranium maximum supply curve.

4.9.9. KEY OUTPUTS

The most relevant outputs of Materials module are as follows:

For the Fossil Fuel submodule:

- Relative RURR oil/ coal/ natural gas
- Oil /coal/ natural gas reserves to production ratio
- Oil/ coal/ natural gas consumption per capita
- Estimated oil/ coal/ natural gas price
- Total oil/ coal/ natural gas extracted
- Oil/ coal/ natural gas resources
- Cumulative total oil/ coal/ natural gas extracted

For the Metal submodule (Fe, Ni, Cu; Al):

- Relative RURR Fe/ Ni/ Cu/ Al
- Fe/ Ni/ Cu/ Al reserves to production ratio
- Fe/ Ni/ Cu/ Al consumption per capita
- Fe/ Ni/ Cu/ Al price economy
- Fe/ Ni/ Cu/ Al extraction
- Fe/ Ni/ Cu/ Al market sales
- Fe/ Ni/ Cu/ Al resources
- Fe/ Ni/ Cu/ Al cumulative mining

Material requirements of energy technologies:

- Materials required for new process transformation capacities: solar PV (on ground and rooftop, 5 sub-technologies), wind onshore and wind offshore, as well as the materials required for the grid expansion.
- Materials required for new railway catenaries.
- Materials required for new storage facilities (for electric transport and stationary systems for back-up of electricity system), including auxiliary elements (chargers, additional grids)
- Embodied intensity in materials

Cumulated materials extracted by energy technology compared with current assessed level of reserves and resources.

4.10. THE SOCIETY MODULE

4.10.1. GENERAL DESCRIPTION

This module, together with the Economy module, is responsible of the social dimension. So far, few but relevant outputs are introduced to reflect the social impacts derived from policies. Figure 53 shows the general flow of information in this module. There are no feedbacks to other modules; society, so far, takes information from other modules to build social indicators.

The composition of households is an input for estimating the consumption patterns. Annual variations are gathered from exogenous data and then linked to population (endogenous).

The level of educational attainment is taken as a reference in many demographic studies because it has a significant influence on different social issues, although education is considered a mere social construct (Lentzner, 2010; Lutz & KC, 2010, 2011). Most countries use the percentage of the population having completed a certain level of education to define their educational profile, according to the statistical reference categorization developed by UNESCO, i.e., the International Standard Classification of Education (Commission et al., 2015). This categorization is useful for comparing different educational profiles across countries or regions. It also provides information on the existing labour force according to its level of education. New workforce is endogenously determined by the inertia of government expenditure on this social service, including the net migration in the region, and exogenous assumptions on the gender parity index and years of education by level. All the variables in this module fulfill requirements to achieve two key indicators, i.e., the human development index (HDI) and the GINI coefficient.

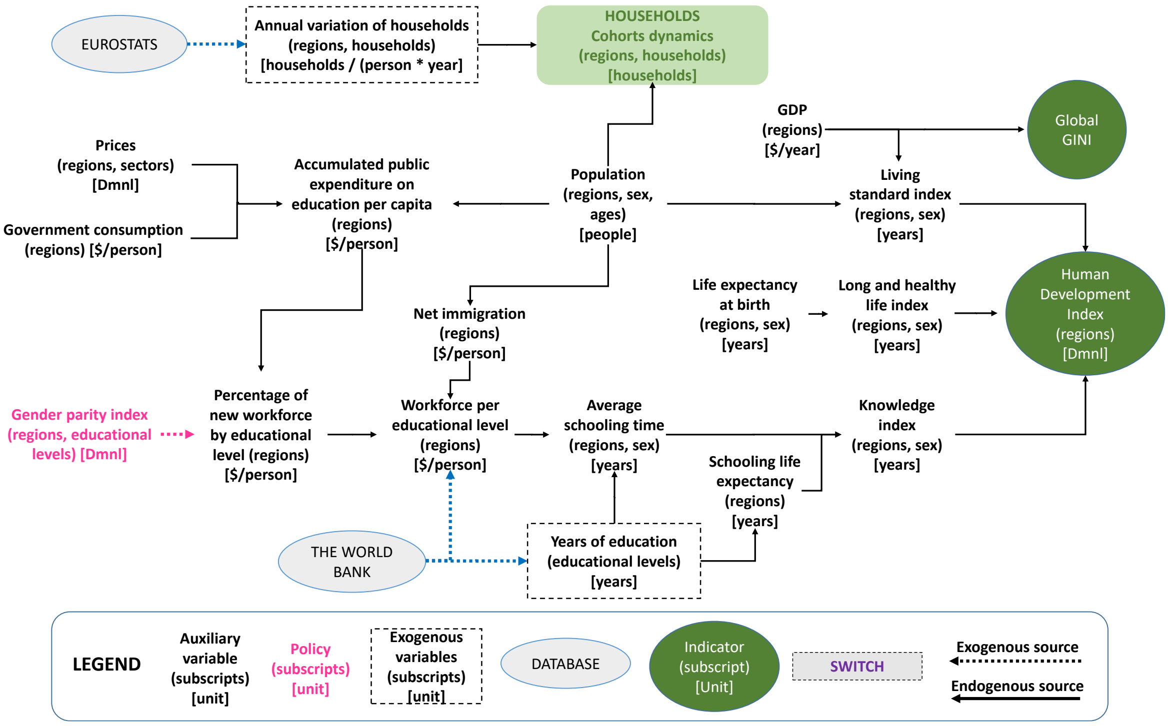


Figure 53: Overview of the Society submodule.

4.10.2. METHODOLOGICAL APPROACH

The *Human Development Index* (HDI) insists on the idea that people and their capabilities should be seen as the guiding criterion when assessing the degree of development of a given country, thus avoiding focusing solely on the rate of economic growth. Therefore, is a summary measure of average achievement in key dimensions of human development: a long and healthy life ($Long\&HeaIdx$), access to knowledge ($KnowledgeIdx$) and having a decent standard of living ($StdLivIdx$).

The HDI (Equation 162) is the geometric mean of normalized indices for each of the three dimensions (United Nations Development Programme., 2020):

$$HDI_t(Reg_i) = \sqrt[3]{KnowledgeIdx_t(Reg_i) \cdot Long\&HeaIdx_t\ (Reg_i) \cdot StdLivIdx_t(Reg_i)}$$

Equation 162

The way in which the function of educational attainment is posited allows us to see a fluid and progressive evolution, since we understand that economic expenditure on education is not the only variable that would affect the level of educational attainment. However, the cumulative average of public expenditure on education over the last 15 years allows to calculate, together with the parameters defined to find the level of educational attainment, the percentage of new labour force for each level of educational attainment. The cumulative average in per capita public expenditure on education is proposed here as a way of breaking down the unrealistic contrasts that would occur in the event of a sudden increase or decrease in public expenditure on education by governments. On the other hand, it is proposed that this average should include data from the previous 15 years, since this is the age that is usually used as a reference when calculating the general level of adult attainment in other models.

With regard to the BETA 0 and BETA 1 parameters, we should point out that they have been calculated for both high and medium levels of education (which later allows us to know the values for low levels of education) and, specifically, for the range of people between 25 and 34 years of age, as we consider that, in addition to finishing their studies, this is the age range where most people enter the labour market.

To this can be added a preceding flow of the gender parity index, which is a socio-economic index usually designed to calculate the relative access of men and women to education. In its simplest form, it is calculated as the ratio of the number of females to the number of males in a given stage of education. It should be noted here that this is a scenario parameter that is set exogenously by the model user, who needs to know that a value below one indicates differences in favor of boys, while a value close to one indicates that parity has been more or less achieved.

All this allows us to know what the percentages of the new labour force are for each level of educational attainment in that particular age group and according to sex. These percentages change, as the level of educational attainment increases or decreases, until they reach certain maximums or minimums. These limits are set by giving a certain margin to the maximum and minimum percentages observed through historical data:

$$.PercH = \beta 0_H + \beta 1_H \cdot \mathrm{APE}$$

Equation 163

$$\boldsymbol{PercH = \beta 0_M + \beta 1_M \cdot APE}$$

Equation 164

$$PercL = \mathrm{Perc}_H - \mathrm{Perc}_M$$

Equation 165

The next step is to know the increase in the labour force. Once we have the percentages previously calculated through the aforementioned parameters, i.e., the new labour force by each level of educational attainment, we have to apply them to the new labour force flows by region and gender. At this point, people leaving the 20-24 age range are counted, as this is the age at which people might be sufficiently educated to enter the labour market for the first time in their lives.

$PPublicExpOnEducationPC$ means public expenditure on education per capita, dynamized within the Economy module (Equation 166). Public expenditure is taken instead of total expenditure (public and private) because of the limitation of historical data, which have been necessary to calculate the parameters of the function of percentages of new workers by educational level.

$$PublicExpOnEducationPC = \frac{\frac{GovernmentConsumptionPurhaserPrice(Reg_i) * 100}{Price\_output(Reg_i)}}{Population(Reg_i)}$$

Equation 166

The workforce in the region is structured by educational level (EL) and follows the next main equations (Equation 167). $ParityIdx$ is the parity index, a socio-economic index to help calculate the access of men and women to education. It is set exogenously by the user of the model. It is calculated as quotient of the number of females by the number of males enrolled in a given stage of education.

$$\frac{Workforce\ (Reg_i, Sex_j, EL_i)}{dt} = \frac{P5y_t(Reg_i, Sex_j, Age_{20-24}) - P5y_t(Reg_i, Sex_j, Age_{60-64}) - deaths_t(Reg_i, Sex_j, Age_{k:\,25-64})}{5} \cdot f(ParityIdx, \beta_0 + \beta_1 \cdot PublicExpOnEducationPC)$$

Equation 167

4.10.3. DATA SOURCES

Eurostat's micro-data have been used to construct the household approach. For education and indicators, the following databases have been used as reference: UIS Stats, OECD Statistics, Barro-Lee Dataset and World Bank Database. All of them have provided the necessary historical data to complete the model.

4.10.4. IMPLEMENTED POLICIES

The Society module contains a set of policies that directly affects to the variables set out above. A user can customize and apply three predefined policies:

- Household composition in Europe-27 (EU27). Value for the switch scenario for household's composition for EU27 countries. 4 options for the evolution of the ratio of households per 100 people over time available from the statistical analysis of past data.
- Number of people per household in non-EU27. Variation over time of average number of people per household for non-EU regions.
- Gender parity in education. Socio-economic index designed to calculate the relative access of men and women to education. This index is released by UNESCO. In its simplest form, it is calculated as the quotient of the number of females by the number of males enrolled in a given stage of education (primary, secondary, etc.). A GPI equal to one signifies equality between males and females. A GPI less than one is an indication that gender parity favors males while a GPI greater than one indicates gender parity that favors females. The closer a GPI is to one, the closer a country is to achieving equality of access between males and females. This is a policy target in WILIAM to be set for a final year per region.

4.10.5. KEY OUTPUTS

The main outputs of this module are final indicators:

- Human Development Index and its independent terms (knowledge index, long and healthy life index, and the standard of living index).
- GINI coefficient based on the GDPpc.
- Gender parity index.
- Workforce per educational level.
- Accumulated public expenditure on education per capita. This output indicates the effort of the government regarding learning and training abilities in population.

4.11. THE WILIAM SCENARIOS

The WILIAM scenarios are defined through the following basic elements: Storylines/narratives for broad socioeconomic trends that could be followed by societies in the future, overall goals acting as “benchmarks” to gauge whether a simulated scenario leads to a desirable solution to the identified problems and specific policy objectives, which are implemented through concrete policy measures. This section elaborates the selection process of the basic elements for scenario definition.

4.11.1. STORYLINES

The set of storylines includes two baseline storylines based on the Shared Socioeconomic Pathways (SSP) - SSP2 “Middle of the road” and SSP3 “Regional rivalry - A rocky road” and three policy-action storylines - Green Growth, Green Deal and Post-growth (Figure 54). SSP2 and SSP3 were selected as baseline storylines because **SSP2** represents a trajectory of the continuation of current trends by definition, whereas **SSP3** represents a future which includes inequality issues (between and within regions), as well as the possibility of the continuation of the reliance on fossil-fuels. For policy-action storylines, Green Growth and Post-growth are taken as extremes and Green Deal as an intermediary storyline. On one hand we have ***Green Growth***, based on market incentives and trickle-down hypothesis (globalization, firm concentration towards achieving economies of scale, quite strict public balance rules, etc.), and ***Post-growth*** based on reducing the role of markets and replacing it with local and government initiatives (re-localization, local economy and small firms, redistribution, etc.) on the other hand. Between these two is the storyline ***Green Deal***, which does not rely exclusively on market mechanisms for the ecological transition, but rather calls for relevant government interventions to prevent increasing inequality and unemployment. It might recall the concept of "inclusive capitalism", in which public reforms aim to enhance the inclusion of all into markets (labour, credit) or at least to avoid exclusion.

Table 8 below summarizes the three proposed storylines. Two goals are crucial for "sustainability": planetary boundaries and prosperity to all (within and between countries). These environmental and socio-economic goals fit within the three pillars of sustainability: people, planet and prosperity. Current climate actions are mainly concentrated at reducing the environmental impact of our societies without changing the political-economic framework (Green Growth). However, institutions recognize that environmental issues are already generating severe social problems and that environmental policies could also negatively affect some social indicators. This reason leads us to think of an alternative to green growth by including radical social policies (Green Deal) that aim at compensating for the negative social side-effects. Finally, Post-growth departs from the assumption that social and environmental goals can only be attained jointly: i) respecting the planetary boundaries entails a reduction of the scale of the economy (especially in high-income countries) and a change in behaviour cannot be promoted by market incentives only, and ii) reducing inequality and avoiding social exclusion requires universal policies and not only specific (technocratic) solutions.

Table 8: Environmental and socio-economic goals in the three proposed storylines. These general goals are considered when setting the overall goals for each scenario to be run in WILIAM.

Goals	Green Growth	Green Deal	Post-growth
Environmental goals	Prioritized - through markets and innovation	Prioritized - through markets, innovation, regulation, and large public expenditures	Prioritized - through transformation of economy, regulation, production slowdown and behavioural change
Socio-economic goals	Not an objective per se, trickle-down (as in baselines)	Specific inclusion policies	Universal social policies

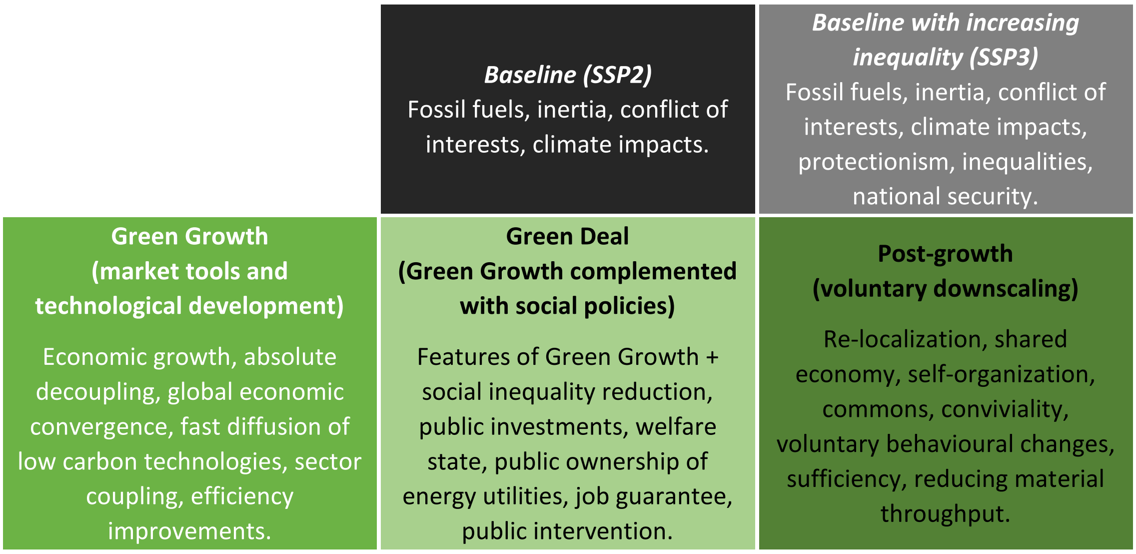


Figure 54: Storylines modelled in LOCOMOTION with main keywords-features.

4.11.2. OVERALL GOALS

Each LOCOMOTION storyline has economic, environmental and socioeconomic goals. These storyline goals (drivers) are used to determine exogenous inputs to the model so as to ensure that the storyline is properly translated into a scenario. WILIAM also distinguishes between overall goals identified as driving a given storyline (storyline goals) and those which the model users may want to check if a given storyline are able to fulfil (expanded overall goals). An example of a storyline goal is "(absolute) decoupling of environmental pressures from economic growth" in the Green Growth storyline and scenario. An example of an expanded overall goal under the Green Growth story is "Universal access to health care/social security/education" or "Gender equality/Non-discrimination of women and girls".

Both baseline storylines SSP2 and SSP3 have the economic goal of GDPpc growth but lack explicit environmental goal. Notably, SSP3 has the socioeconomic goal of maintaining national sovereignty due to the logic of the narrative.

While Green growth and Green deal have GDPpc growth as an economic goal, Post-growth is driven by the idea of economic re-localization. Environmental goals are prioritized in all storylines, albeit through different means. Finally, Green growth does not have a socioeconomic goal, while Green deal and Post-growth have specific inclusion policies and universal social policies as drivers respectively.

The list of WILIAM overall goals (for more details, see (Luzzati T. et al., 2021)) has been linked to storylines and storyline goals (drivers) specifically. For example, the overall goal to use land sustainably is linked to storyline goals (drivers) in all three storylines, whereas the overall goal of universal access to education is only linked to storyline goals (drivers) in Green Deal and Post-growth. Moreover, the overall goals of minimizing migration and gender equality are only linked to expanded overall goals and not to the storyline drivers.

4.11.3. POLICY OBJECTIVES AND POLICY MEASURES

Another step necessary to a proper definition of policy scenarios, is grouping the policy measures selected in (Böck et al., 2020)(over 1300) under 129 policy objectives which then have been ranked on the base of modelers' guess, stakeholders' sentiment, and analytical feasibility (within WILIAM). The prioritization procedure is explained in detail in (Luzzati T. et al., 2021).

The main outcome of the ranking analysis is summarized in the Table 9.

Table 9: Ranking of the selected policy objectives.

#	Policy objective	Rank
34.	Energy efficiency: Transport	1
80.	Land use: Afforestation	2
126.	Unemployment reduction	3
88.	Modal shift: electric mobility	4
57.	GHG emissions: Food	5
84.	Land use: Regenerative Agriculture	6
32.	Energy efficiency: public buildings	7
69.	Heat generation: Solar	8
104.	Resources: Land management	9
4.	Climate change adaptation	10
90.	Modal shift: railway	11
73.	Improved LULUCF capacity	12
66.	Heat generation: district heating	13
22.	Electrification	14
63.	Green growth	15
82.	Land use: Increase organic agriculture	16
67.	Heat generation: Geothermal	17
78.	Labour reform: salary	18
76.	Labour reform: productivity	19
109.	Resources: Reduction of extractive activities	20
61.	GHG emissions: Livestock	21
101.	Resources: Increased use of biogas	21
129.	Work-Life Balance	22
44.	Equitable and resource efficient governance	23
35.	Energy efficiency: transport - airplanes	25
87.	Modal shift: bikes	25
89.	Modal shift: public transport	25
107.	Resources: Recycling - households	28
36.	Energy efficiency: Transport - Passenger cars	29
38.	Energy efficiency: transport - trucks	29
43.	Energy transmission infrastructure	31
17.	Electricity generation: RES	33
85.	Land use: Soil protection/fertility	33
77.	Labour reform: retirement	34
49.	Fiscal reform: ecological	35

125.	Transport: Reduced transport volume - people	90
110.	Resources: Reduction of food waste	91
7.	Consumer protection	92
46.	Family Planning	94
96.	Regionalization of production	95
108.	Resources: Recycling - Industry	95
2.	Alternative structure of economy	97
86.	Minimize the rebound effect	97
121.	Transport: Emissions reduction	97
27.	Energy efficiency: Buildings	100
92.	Modal shift: Ships	100
123.	Transport: Reduced transport volume - flights	100
24.	Energy efficiency: Agriculture	102
47.	Financial reform: markets	102
52.	Fuel shift: Green H2 production & use	102
3.	Animal health	105
12.	Electricity generation: in-stream hydro	105
56.	GHG emissions: F-gases	107
30.	Energy efficiency: Industry	108
74.	Increase economic stability	109
75.	Increased public property ownership	110
37.	Energy efficiency: Transport - Public transport	111
94.	Reduce overall debt	111
100.	Resources: Increased use of alternative cement	113
102.	Resources: Increased use of biomass	113
13.	Electricity generation: nuclear	115
122.	Transport: Infrastructure	115
117.	Stabilize population growth	117
91.	Modal Shift: ride sharing	118
79.	Labour reform: working hours	119
60.	GHG emissions: Industry	120
5.	Climate change mitigation: Demand side	121
106.	Resources: New materials and technologies	122
53.	Fuel shift: Industry	123
112.	Resources: Treatment of landfill gas	124
26.	Energy efficiency: Behaviour building users	125
71.	Improve education for girls	125
45.	Ethical Finance	126
48.	Financial reforms: central bank	128
59.	GHG emissions: Improved carbon pricing	129

4.12. VERSIONS AND INTERFACES OF THE WILIAM MODEL

The WILIAM model is accessible to the public in 5 different flavours:

1. **WILIAM**: Vensim® version of the WILIAM model. It requires users to download and install the Vensim® Model Reader software (Vensim Model Reader – Ventana Systems), which is a proprietary but free to use software by VENTANA systems.inc.
2. **pywiliam**: fully free and open-source version of the WILIAM model written in Python.
3. **The Model Analyzer**: desktop application that allows configuring scenarios, running the WILIAM model and plotting simulation results. Its main target audience are stakeholders and experts.
4. **The Model Explorer**: web application that allows users to modify a few scenario parameters and view WILIAM outputs in real time. This web app is addressed at the general public.

5. **The Global Sustainability Crossroads II Game**: gamification tool based on the WILIAM model, addressed at high school and pre-graduate students.

A description of each of these products and links to access or download them or their respective codes are available through the project website (https://www.locomotion-h2020.eu/), from the *LOCOMOTION model* menu.

4.13. MODEL TRANSPARENCY

The development of a shared database is a key part of the project, as it aims to provide a reliable and transparent mean that gives an overview on the Symbols used in the WILIAM model, their metadata and their continuous update, both for the modellers involved in the development of the WILIAM model, as well as for external users. The shared database also aims to provide a mean for the organisation of information for the dissemination of project outputs.

The shared database and open database management system developed in the LOCOMOTION project consists of three different parts:

- The first part of the database, which is called *Data Dictionary*, includes information on the Symbols developed within the project (variables, historical data, constants, parameters, scenario parameters, etc.) and used in the models and information related to their metadata, as well as acronyms, semantic rules, etc. A specific protocol has been defined for the Data Dictionary in order to validate modifications based on the roles of users. The purpose of the Data Dictionary is to enhance the transparency of the model by disclosing all the Symbols and data sources used.
- The second and third part of the database of the narrative description of simulated scenarios and the main output variables of WILIAM include information on the narrative description of selected storylines and selected simulated scenarios and the necessary details to access the inputs used for these scenarios, as well as, the future projections of the main output variables of the model for selected simulated scenarios. The purpose of the database of the narrative description of simulated scenarios and the main output variables of WILIAM is to be used as an instrument for dissemination of the results of WILIAM.

4.13.1. DATA DICTIONARY

4.13.1.1. ACCESS TO THE DATA DICTIONARY AND MAIN PAGE

The Data Dictionary web application is currently hosted in CRES with URL: http://www.cres.gr/LOCOMOTION_client. By the end of the project the Data Dictionary will be accessible in the LOCOMOTION project website with URL: https://www.locomotion-h2020.eu/locomotion-models/database/.

In the main page, the user has two options in order to proceed:

1) The first is to proceed to the Public View as a non-registered user. This option is available for external users to the LOCOMOTION project in order to provide them with enough information with respect to the WILIAM model and enhance the transparency of the model by disclosing information on all the Symbols and data sources used. The subsequent chapters of this Section focus on the features and functionalities of the Data Dictionary available for non-registered users.
2) The second is to Log in with their credentials. This option is only available for the LOCOMOTION modelling groups in order to allow the creation, maintenance and support of the modelling groups in the use of data during the development of the WILIAM model. More information on

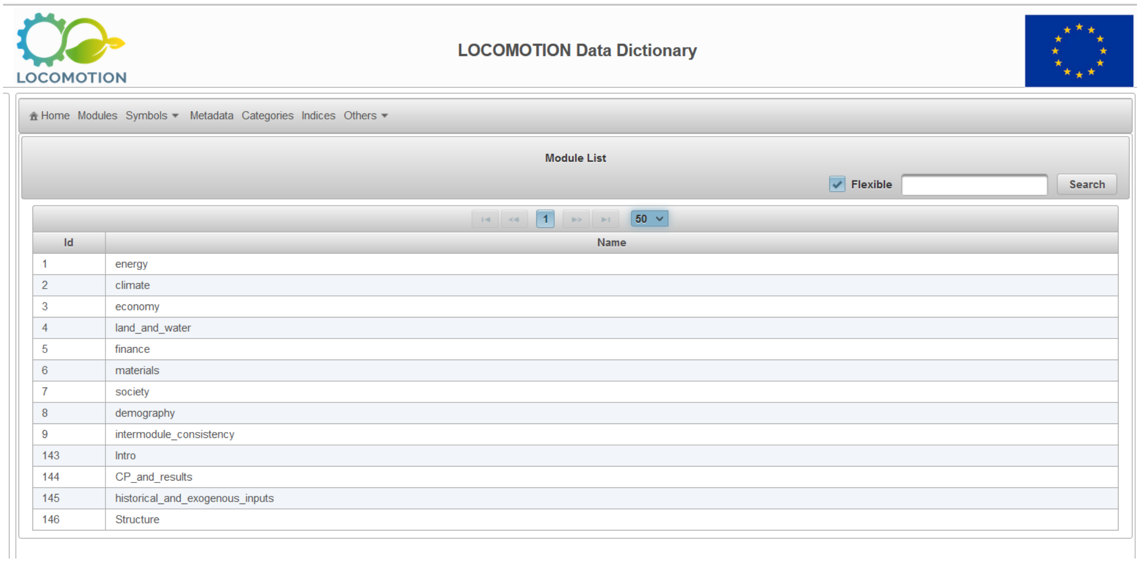
the features and functionalities of the Data Dictionary by type of registered user is available in the document “Shared database and Open DBMS manual”.

4.13.1.2. MAIN ENTITIES OF THE DATA DICTIONARY

The main entities included in the Data Dictionary are presented below.

4.13.1.2.1. MODULES

The WILIAM model consists of 8 different modules, as seen in Figure 55. Each Symbol has a main module, and it can be used in more than one module.



Id	Name
1	energy
2	climate
3	economy
4	land_and_water
5	finance
6	materials
7	society
8	demography
9	intermodule_consistency
143	Intro
144	CP_and_results
145	historical_and_exogenous_inputs
146	Structure

Figure 55: Module List main page.

4.13.1.2.2. SYMBOLS

A large set of Symbols are introduced in the WILIAM model. The type of Symbols includes: Historical Data Series, Constants, Scenario Parameters, Variables, Indices, Index Cases, Switches, Model Parameters, Validity rules. The definitions and examples for each type of Symbol are available in the Data Dictionary (Others/Project Type of Values). For a Symbol to become available in the Data Dictionary Public View, it must first be validated.

4.13.1.2.3. METADATA

Metadata is data that provide descriptive information about the content and production of the data in order to facilitate the use and interpretation of the supplied data.

In the LOCOMOTION project, the information relating to structural metadata, which describe or retrieve statistical data, such as Symbol names, dictionaries, dataset technical descriptions, data formats, keywords in order to find data etc. are made available through the Data Dictionary. Reference or descriptive metadata, which describe the content, context and quality of the statistical data, such as description, methodology for data collection etc., are presented based on a standardized format, used by Eurostat, called ESMS (Euro SDMX Metadata Structure). Considering the goal and the requirements of the project, a more simplified version has been developed based on the elements that can be adapted from standard metadata schemes. The guidelines of the type of information recorded as metadata elements and their format and content rules developed allow for higher consistency (Anderson, 2007; Eurostat, 2015, 2019; Nations, 1995).

Initially, the documentation of metadata for the datasets that have been collected and compiled within the project has been performed in a spreadsheet file. However, in the shared database and database management system, metadata are available for Symbols. The relationship between Symbols and Metadata is a one-to-many relationship, while metadata records can exist without being associated with a Symbol and vice versa. A Symbol can only be connected to one metadata record and a metadata record can belong to several Symbols. Metadata can be uploaded in the database without having information on which Symbol (or Symbols) they are associated with. In order to facilitate the process, the metadata template that has been created for the collected and compiled datasets has been incorporated in the shared database and database management system.

Overall, 38 metadata fields have been created, a detailed description of which is available in Table 10. More detailed information on the metadata fields and their properties is available in Deliverables 2.2 and 2.7 of LOCOMOTION project.

Table 10: Metadata fields in WILIAM.

	Metadata field	Metadata field description
1	Dataset/Symbol ID	Identifier assigned to a single dataset/Symbol.
2	Dataset/Symbol name	A name given to the dataset, i.e., any organized collection of data or Symbol.
3	Main module	The main module that has requested the dataset or the module primarily responsible for the Symbol.
4	Secondary module(s)[20]	Secondary module(s) that have requested or intend to use the dataset or the Symbol.
5	Project type of value [20]	Type of value of the dataset/Symbol.
6	Keywords [20]	Insert keywords separated by comma to assist the search of Symbols.
7	Values read from	Indication of the specific location where the dataset/Symbol values can be found in the model Input files.
8	Unit of measure	Unit in which the data values of the dataset/Symbol are expressed.
9	Statistical concepts, definitions and methodology	Main characteristics of the dataset/Symbol. The method used and operations performed to gather and compile data.
10	Source data	Data source(s) where the data have been obtained from.
11	Last source update	Date of last update of the content of the data in the original source.
12	Link	Link to the original data source(s).
13	Confidentiality/License	Property of data indicating whether they are subject to dissemination restrictions.
14	Data collection date	Date of collection of the data.
15	Time coverage	The length of time for which data have been collected.
16	Metadata update [20]	Date of last update of the content of the metadata.
17	Contact organisation	The name of the organisation that compiled the data/Symbol or metadata.

[20] Only applicable in the metadata fields for Symbols included in the Data Dictionary and not in the metadata fields created in relation to collected datasets.

	Metadata field	Metadata field description
18	Comments	Supplementary descriptive text and explanation of any special circumstances or other information that may affect the interpretation of the data.
	Reliability	Reliability index of the data sources used in data collection.
19	Reliability of data sources	Result of the survey to identify whether a dataset is suited to be used to generate new data for modelling purposes based on credibility of the author, objectivity/bias, accuracy and reliability and currency and timeliness.
	Uncertainty [20]	Values of the statistical descriptors considered in the uncertainty analysis.
20	Type of input data	Method used to generate data.
	1) Set of values	The case where parameters are directly derived from data, e.g., data from external databases, from literature, from own experimental data of the module developers.
21	Sample mean	Average of all values obtained with the following expression: $$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i = \frac{x_1 + x_2 + \ldots + x_N}{N}$$ where $\{x_1, x_2, \ldots, x_N\}$ are the observed data and N is the number of data points.
22	Sample standard deviation	Standard measure describing the spread of the data obtained with following expression: $$s = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2}$$ where $\{x_1, x_2, \ldots, x_N\}$ are the observed data, N is the number of data points and $\bar{x}$ is the sample mean.
23	Sample maximum	Maximum value of all sampled values.
24	Sample minimum	Minimum value of all sampled values.
25	Most-likely maximum	Maximum value of the range of likely values for a certain parameter, estimated using previous knowledge of the variable at stake.
26	Most-likely minimum	Minimum value of the range of likely values for a certain parameter, estimated using previous knowledge of the variable at stake.
27	Physical top boundary	Top boundary for values that are not physically possible for certain parameters (e.g., a parameter can only take on positive values (x>0), fractions that must be between 0 and 1).
28	Physical lower boundary	Lower boundary for values that are not physically possible for certain parameters (e.g., a parameter can only take on positive values (x>0), fractions that must be between 0 and 1).
	2) Model/regression	The case when the module developers are using models and performing regression analysis. Differentiation between the use of external models to WILIAM which are used to provide inputs into WILIAM (the parameter inputs considered could be either

	Metadata field	Metadata field description
		estimated model coefficients or output variables of the external model) and performing regressions on the module itself (the estimated model coefficients considered as parameter inputs).
29	Best estimate	In the case of an estimated model coefficient, the best estimate obtained in the regression. In the case of a regression model output variable, its best estimate obtained in the regression.
30	Standard error	In the case of an estimated model coefficient, the standard error is obtained from the regression results (square root of the diagonal elements of the variance-covariance matrix). In the case of a regression model output the information it is contained in the information used for computing associated confidence intervals.
	3) Expert knowledge	The case where input parameters need to be obtained from expert knowledge by the modellers and module developers.
	a) Schematic pathway	In the case less knowledge is available.
31	Lower value	A lower bound considered in a schematic representation of uncertainty.
32	Upper value	A lower bound considered in a schematic representation of uncertainty.
33	Most-likely value	An upper bound considered in a schematic representation of uncertainty.
	b) Statistical pathway	In the case more knowledge is available.
34	Best estimate	A most likely value considered in a schematic representation of uncertainty.
35	Standard deviation	In a statistical representation of uncertainty, the best estimate, i.e., expected value of a distribution.
	4) No statistical descriptors added	No statistical descriptors are included.
36	No statistical descriptors - explanation	Explanation of the reasons why statistical descriptors could not be provided.
37	Qualitative appraisal of the uncertainty	Qualitative appraisal of the uncertainty of the parameter.
38	Probability distribution	Name of the distribution and the coefficient values of the parameters in case the distribution is other than uniform, beta-PERT/triangular, normal, log-normal.

4.13.1.2.4. CATEGORIES

Categories are implemented in a tree structure of a maximum of three levels (this means that a category can have up to two sub-categories). A Symbol can belong to only one category (or sub-category). The main page of categories can be seen in Figure 56.

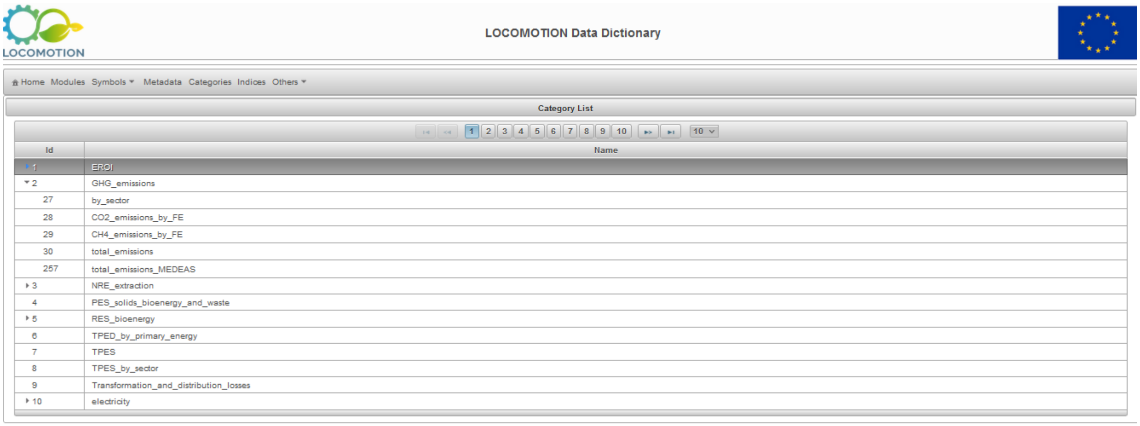


Figure 56: Category List main page.

4.13.1.2.5. INDICES

Some Symbols are declared as indexed and are therefore associated with Indices. An Index is created along with its value or values. Users can view existing Indices. Users can filter indices with their name, doing a flexible or not search. The main page of the Index list can be seen in in Figure 57.

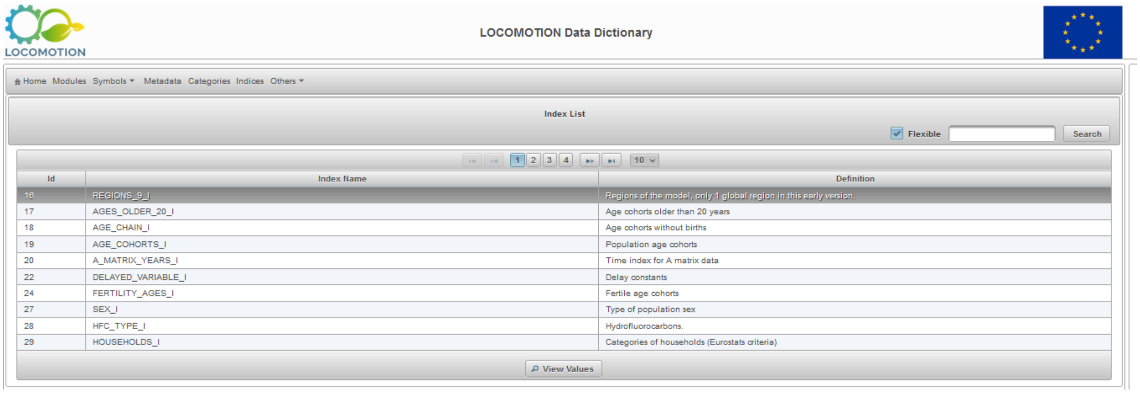


Figure 57: Index List main page.

4.13.1.2.6. OTHER ENTITIES

Other entities include:

- Acronyms, Adjectives, Semantic Rules, Unit Systems: These entities are created to facilitate the tasks of modelling and programming and the understanding of the project by external users, i.e., people outside of the LOCOMOTION project.
- Project Type of Values, Programming Language Symbol Types: Each Symbol is characterized by a type that its value has in the project (variable, constant, time-series data etc.) and this type is matched with a corresponding programming language (Vensim® for LOCOMOTION) Symbol type.
- Sections: This entity includes the sections of the modules.

4.13.1.3. FUNCTIONALITIES OF THE DATA DICTIONARY FOR NON-REGISTERED USERS

Public View of the Data Dictionary is available for non-registered users. This option is available for external users to the LOCOMOTION project in order to provide them with enough information with respect to the WILIAM model and enhance the transparency of the model by disclosing information on all the Symbols and data sources used. Non-registered users can view almost all the information in the Data Dictionary without any personal information about the registered users of the system and without the permission to add or modify any entity object. A non-registered user can navigate through the Data Dictionary web application as described below.

4.13.1.3.1. MAIN PAGE

The main page of the Data Dictionary, the available Symbol actions and other options for a non-registered user are presented in Figure 58, Figure 59 and Figure 60.

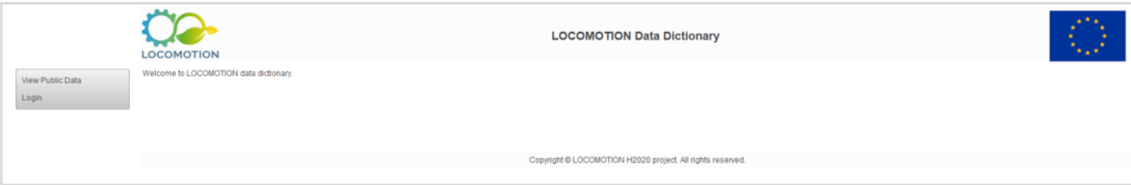


Figure 58: Main page of the LOCOMOTION Data Dictionary for non-registered users.

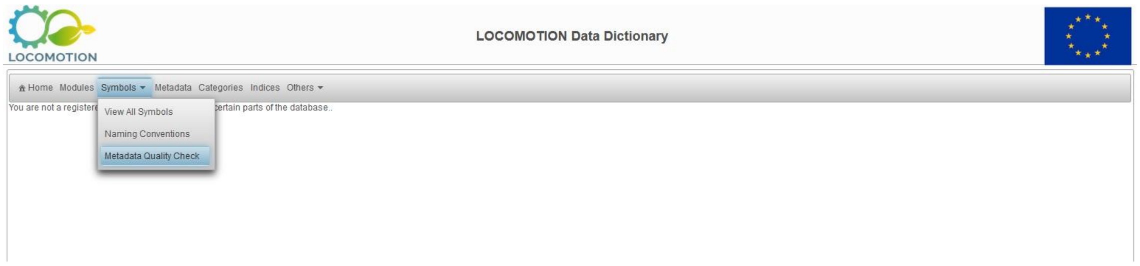


Figure 59: Symbol actions available for non-registered users.

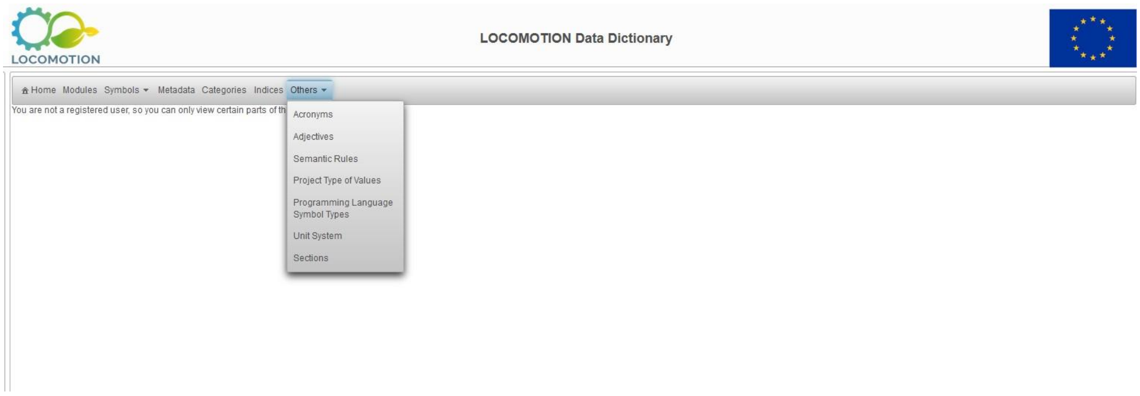


Figure 60: Other options available for non-registered users.

4.13.1.3.2. SYMBOLS

VIEW SYMBOLS

The user can view and obtain information, such as Name, Definition, Category, Main module and Metadata, for all the validated Symbols associated with the WILIAM model, as it can be seen in Figure 61. The user can view all the Symbols of the WILIAM model through:

Menu/ Symbols/ View All Symbols

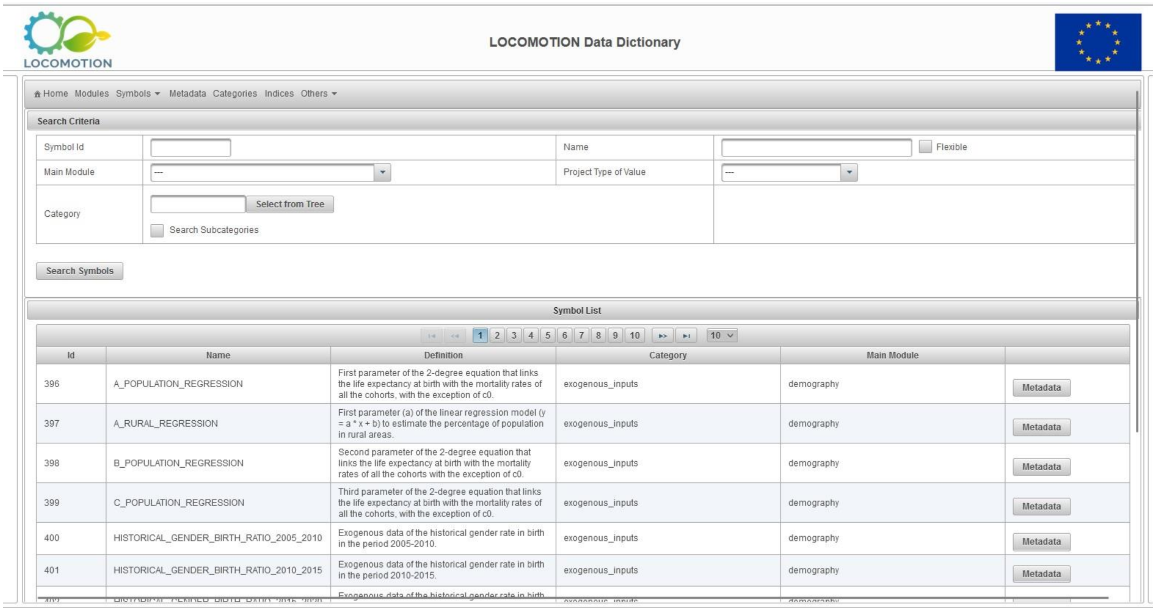


Figure 61: View All Symbols page.

In this page the user has several options:

Search for a Symbol

The user can use search criteria to filter the Symbol. These criteria include the Symbol ID, the Symbol name, the main module of the Symbol, the Symbol category and the Project Type of Value of the Symbol. In case the search is performed based on the Symbol Name, the search can be flexible or not, i.e., the Symbol Name must match exactly or partially the search text. An example of a flexible Symbol Name search is presented in Figure 62.

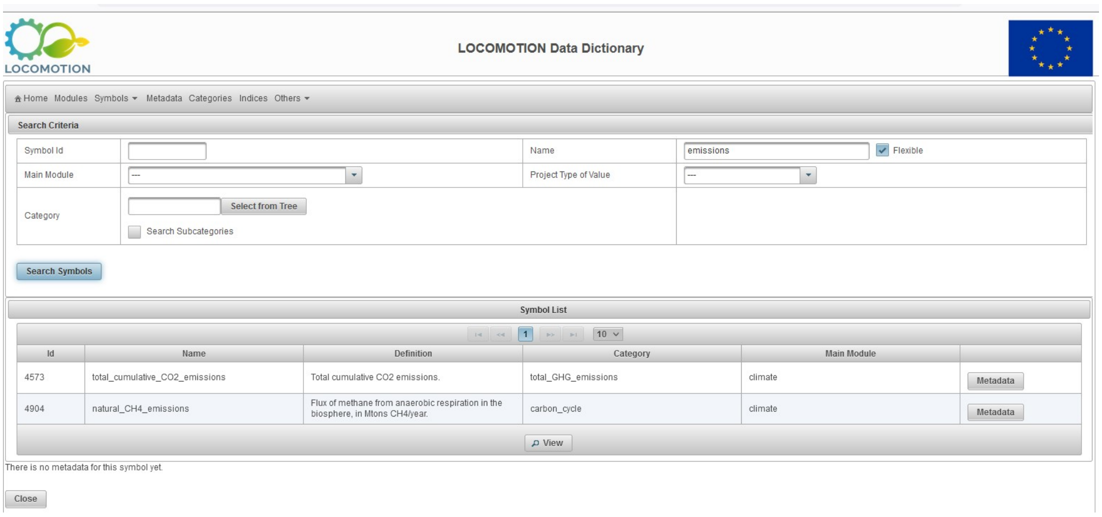


Figure 62: Example of a flexible Symbol Name search.

View Symbol

The user can **view all the details of a Symbol** (Figure 63). These include Symbol attributes, Input/Output attributes, Metadata, Secondary modules in which it is used and associated Indices through:

Menu/ Symbols/ View All Symbols: Select a Symbol in list and press ”View”

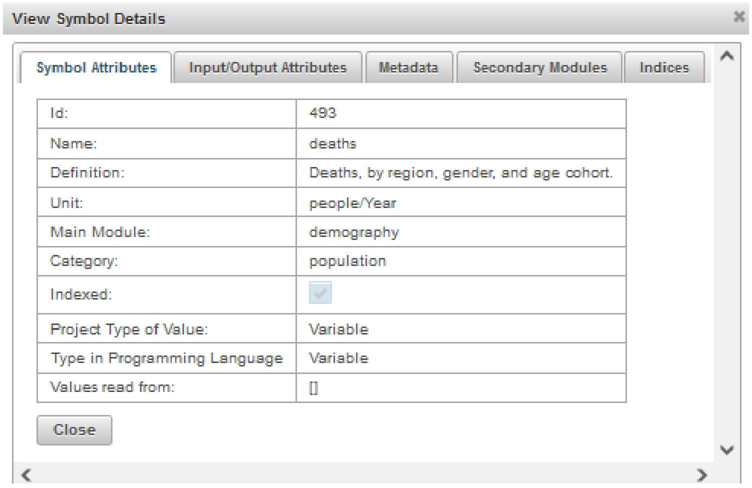


Figure 63: View Symbol details.

NAMING CONVENTIONS

The user can view detailed instructions on the Naming Conventions of the Symbols though:

Menu/ Symbols/ Naming Conventions

This section includes information on the sources that have been considered to define naming conventions, the types of Symbols, Naming Convention rules and Automatic checking of naming convention rules.

METADATA QUALITY CHECK

The user can view the results of the Metadata Quality Checks through:

Menu/ Symbols/ Metadata Quality Checks

This section includes information on the Metadata Quality Checks that are performed by date, Asset (IndexValue, IndexDefinition, Metadata, Symbol) and Expectation Suite by level of severity (Warning, Critical, Error). By clicking on each asset and Expectation Suite users can see the overview of the quality check and the results by type of expectation. The metadata quality checks that run in the Data Dictionary by level of severity are presented in detail in (Meta)data quality checks.

4.13.1.3.3. METADATA

View Metadata

The user can view Metadata through:

Menu/ Metadata

This feature allows the user to have a general overview of the Metadata that have been used in the preparation of the Symbols. The user can view all the Metadata details of a dataset through:

Menu/ Metadata: Select a dataset in list and press ”View”

A detailed description of the Metadata fields that have been included is available in section 4.13.1.2.3 Metadata. The user can download the Metadata in several formats, such as xlsx, csv and xml.

4.13.1.3.4. INDICES

View existing Indices

The user can view existing Indices (Figure 64) through:

Menu/ Indices

The user can filter indices with their name, doing a flexible or not search.

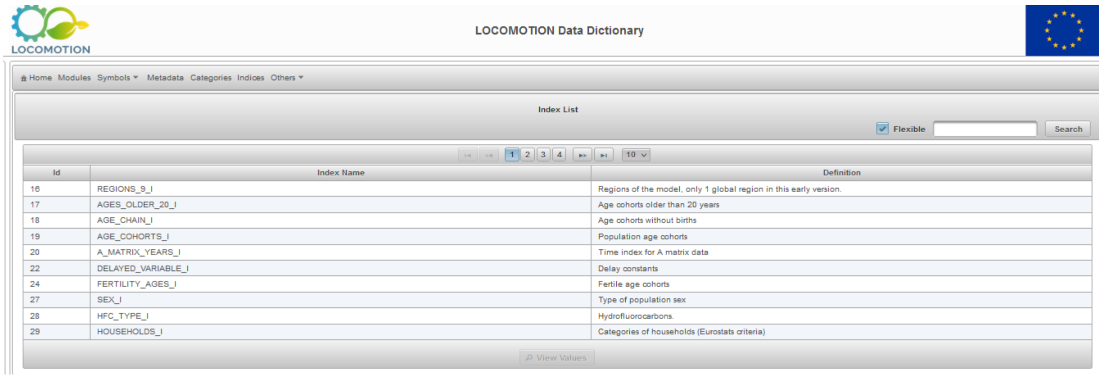


Figure 64: Indices List main page.

The user can view all the details of an Index through:

Menu/ Index: Select an Index in list and press "View Values"

4.13.1.3.5. OTHERS

The user can access other entities of the Data Dictionary through:

Menu/ Other entities

The features of these entities are described below.

ACRONYMS

The user can access the list of acronyms used and their meaning (Figure 65) through:

Menu/ Others/ Acronyms

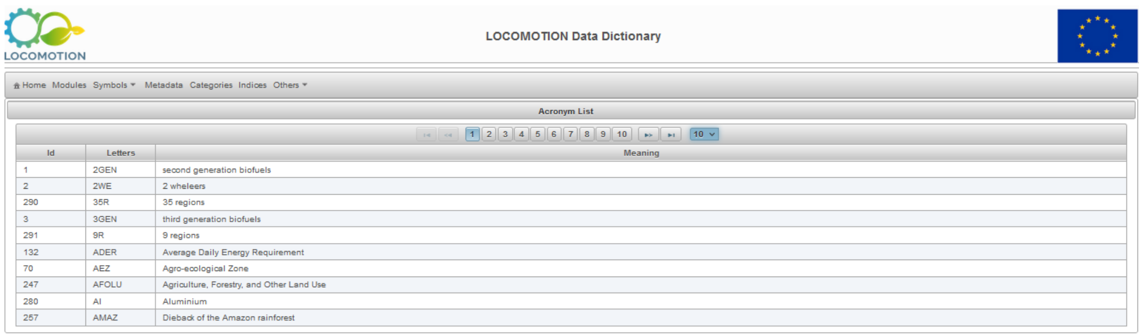
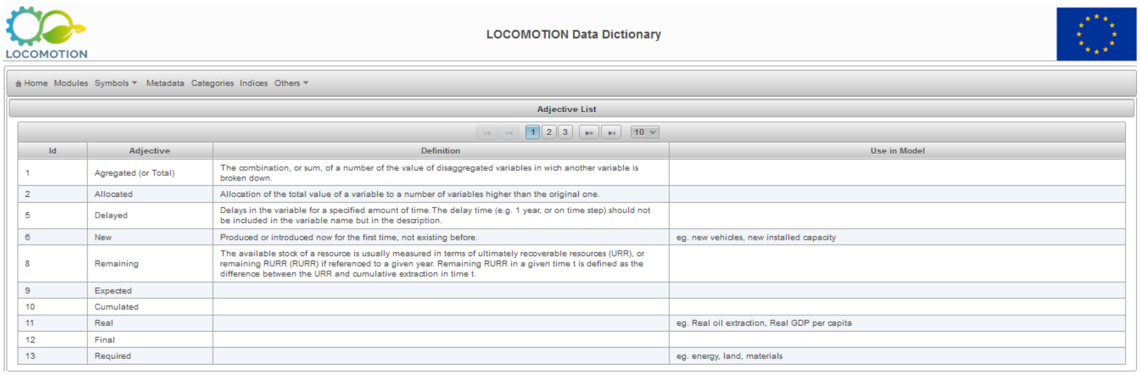


Figure 65: Acronym List main page.

ADJECTIVES

The user can access the list of adjectives used, their definitions and use in the model (Figure 66) through:

Menu/ Others/ Adjectives



LOCOMOTION Data Dictionary

Home Modules Symbols Metadata Categories Indices Others

Adjective List

Id	Adjective	Definition	Use in Model
1	Agregated (or Total)	The combination, or sum, of a number of the value of disaggregated variables in wich another variable is broken down.	
2	Allocated	Allocation of the total value of a variable to a number of variables higher than the original one.	
5	Delayed	Delays in the variable for a specified amount of time.The delay time (e.g. 1 year, or on time step) should not be included in the variable name but in the description.	
6	New	Produced or introduced now for the first time, not existing before.	eg. new vehicles, new installed capacity
8	Remaining	The available stock of a resource is usually measured in terms of ultimately recoverable resources (URR), or remaining RURR (RURR) if referenced to a given year. Remaining RURR in a given time t is defined as the difference between the URR and cumulative extraction in time t.	
9	Expected		
10	Cumulated		
11	Real		eg. Real oil extraction, Real GDP per capita
12	Final		
13	Required		eg. energy, land, materials

Figure 66: Adjective List main page.

SEMANTIC RULES

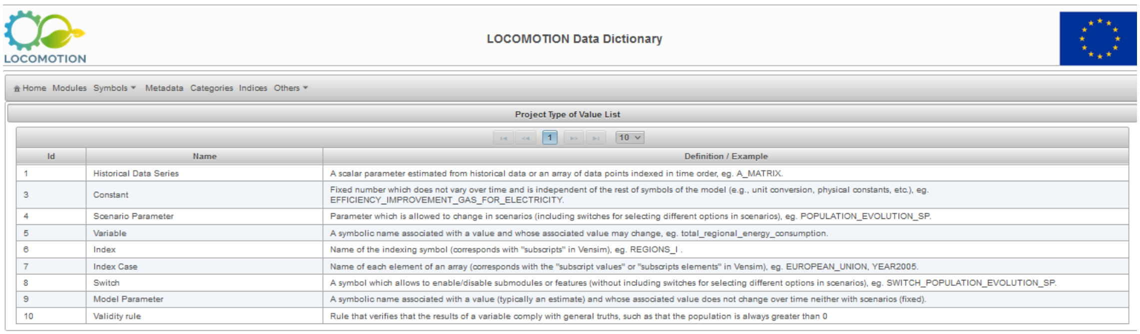
The user can access the list of semantic rules used and their explanation through:

Menu/ Others/ Semantic rules

PROJECT TYPE OF VALUES

The user can access the list of Project Type of Values and their definitions (Figure 67) through:

Menu/ Others/ Project Type of Values



LOCOMOTION Data Dictionary

Home Modules Symbols Metadata Categories Indices Others

Project Type of Value List

Id	Name	Definition / Example
1	Historical Data Series	A scalar parameter estimated from historical data or an array of data points indexed in time order, eg. A_MATRIX.
3	Constant	Fixed number which does not vary over time and is independent of the rest of symbols of the model (e.g., unit conversion, physical constants, etc.), eg. EFFICIENCY_IMPROVEMENT_GAS_FOR_ELECTRICITY.
4	Scenario Parameter	Parameter which is allowed to change in scenarios (including switches for selecting different options in scenarios), eg. POPULATION_EVOLUTION_SP.
5	Variable	A symbolic name associated with a value and whose associated value may change, eg. total_regional_energy_consumption.
6	Index	Name of the indexing symbol (corresponds with "subscripts" in Vensim), eg. REGIONS_I .
7	Index Case	Name of each element of an array (corresponds with the "subscript values" or "subscripts elements" in Vensim), eg. EUROPEAN_UNION, YEAR2005.
8	Switch	A symbol which allows to enable/disable submodules or features (without including switches for selecting different options in scenarios), eg. SWITCH_POPULATION_EVOLUTION_SP.
9	Model Parameter	A symbolic name associated with a value (typically an estimate) and whose associated value does not change over time neither with scenarios (fixed).
10	Validity rule	Rule that verifies that the results of a variable comply with general truths, such as that the population is always greater than 0

Figure 67: Project Type of Value List main page.

PROGRAMMING LANGUAGE SYMBOL TYPES

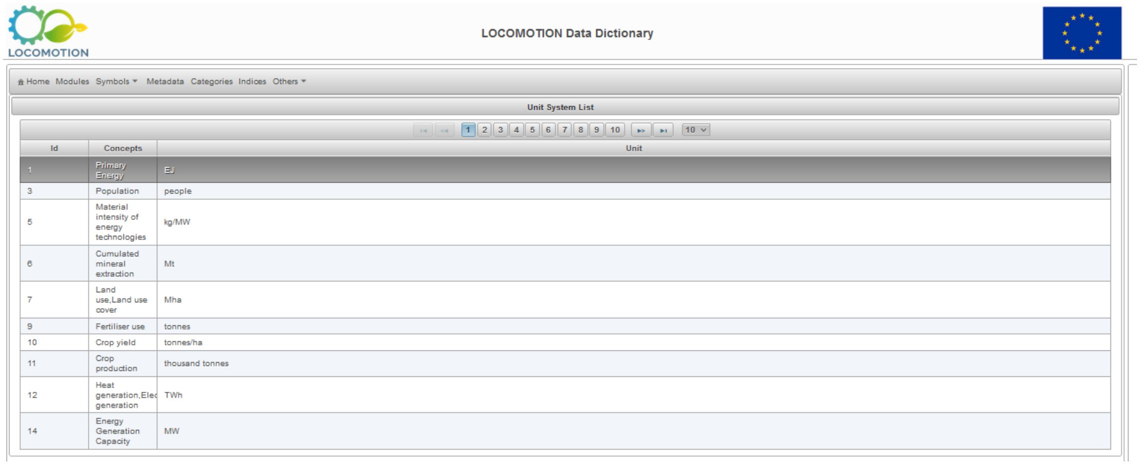
The user can access the list of Programming Language Symbol Types and their definitions through:

Menu/ Others/ Programming Language Symbol Types

UNIT SYSTEM

The user can access the list of Units used (Figure 68) through:

Menu/ Others/ Unit system



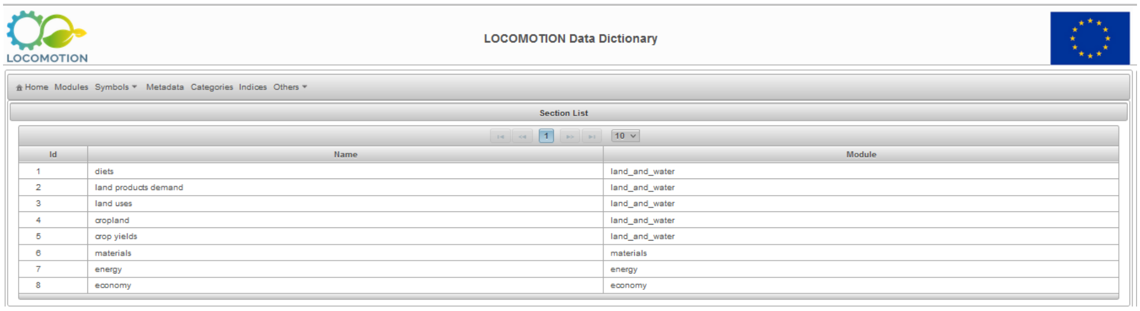
Id	Concepts	Unit
1	Primary Energy	EJ
3	Population	people
5	Material intensity of energy technologies	kg/MW
6	Cumulated mineral extraction	Mt
7	Land use,Land use cover	Mha
9	Fertiliser use	tonnes
10	Crop yield	tonnes/ha
11	Crop production	thousand tonnes
12	Heat generation,Elec generation	TWh
14	Energy Generation Capacity	MW

Figure 68: Unit System List main page.

SECTIONS

The user can access the list of Sections (Figure 69) through:

Menu/ Others/ Sections



Id	Name	Module
1	diets	land_and_water
2	land products demand	land_and_water
3	land uses	land_and_water
4	cropland	land_and_water
5	crop yields	land_and_water
6	materials	materials
7	energy	energy
8	economy	economy

Figure 69: Section List main page.

4.13.2. WILIAM DATABASE OF SELECTED SIMULATED SCENARIOS AND RESULTS

4.13.2.1. ACCESS TO THE WILIAM DATABASE OF SELECTED SIMULATED SCENARIOS AND RESULTS AND MAIN PAGE

The aim of the WILIAM Database of selected simulated scenarios and results is to be used as an instrument for dissemination of the results of WILIAM.

The web application of the Database of the main output variables of WILIAM is currently hosted in CRES with URL: http://www.cres.gr/LOCOMOTION_scenarios. By the end of the project, the Database will be available in the LOCOMOTION project website with URL: https://www.locomotion-h2020.eu/locomotion-models/database/.

In the main page, the user has two options in order to proceed:

1) The first is to proceed to the Public View as a non-registered user. This option is available for external users to the LOCOMOTION project.
2) The second is to Log in with the credentials already available in the Data Dictionary. This option is only available for the LOCOMOTION groups in order to allow the creation, maintenance and editing of the data availability in the Database.

Due to the fact that the main aim of the database is to present selected scenarios and the main output variables for the dissemination of the project results, once the selection is complete, the database will become openly available to all users presenting only the selected WILIAM results. The main page of the WILIAM Database of selected simulated scenarios and results is presented in Figure 70.

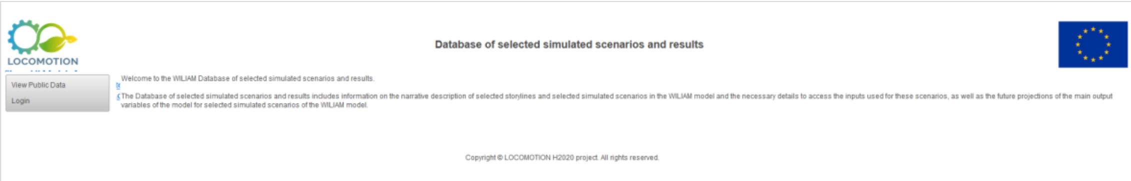


Figure 70: Main page of the WILIAM Database of selected simulated scenarios and results.

4.13.2.2. MAIN ENTITIES OF THE DATABASE OF SELECTED SIMULATED SCENARIOS AND RESULTS

The main entities included in the Database of selected simulated scenarios and results are presented below.

4.13.2.2.1. STORYLINES

This entity includes the different storylines on which modellers are based in order to run different scenarios. The storylines selected in WILIAM and the main keywords/features identifying each storyline, according to Deliverable 8.3 (Markovsa, 2021), are presented below:

- SSP 2 “Middle of the road” (O’Neill et al., 2017)
- Green Growth (GG)
- Green Deal (GD)
- Post-growth (PG)

4.13.2.2.2. SCENARIOS

This entity includes a selection of simulated scenarios in the WILIAM model with their description. Each scenario is based on one storyline, and a storyline can lead to more than one different scenarios.

4.13.2.2.3. SCENARIO POLICY MEASURES AND HYPOTHESES

Each simulated scenario includes some input parameters in the form of predefined policy measures and hypotheses. These parameters are Symbols already stored in the Data Dictionary, having their “Project Type of Value” property set as “Scenario Parameter”. These scenario policy measure and hypothesis entities are available in the database for each scenario.

4.13.2.2.4. SCENARIO OUTPUTS – FUTURE PROJECTIONS

Each simulated scenario leads to some results. These results are Symbols already stored in the Data Dictionary, having their “Project Type of Value” property set as “Variable”. Each Scenario Output entity represents a model simulation obtained at a specific time. Each Future Projection Entity represents the result for Symbols of a specific simulation along with their indices (e.g., regions), if they are indexed.

4.13.2.2.5. FUTURE PROJECTIONS VALUES

Each scenario output is stored as a timeline series representing the future projection values for the output variables. The future projections of the main output variables are available as tables and plots.

4.13.2.3. FUNCTIONALITIES OF THE DATABASE OF SELECTED SIMULATED SCENARIOS AND RESULTS FOR NON-REGISTERED USERS

A non-registered user can access the menu of the WILIAM Database of selected simulated scenarios and results through "View public data". The overview of the Home page and its main entities is provided in Figure 71.

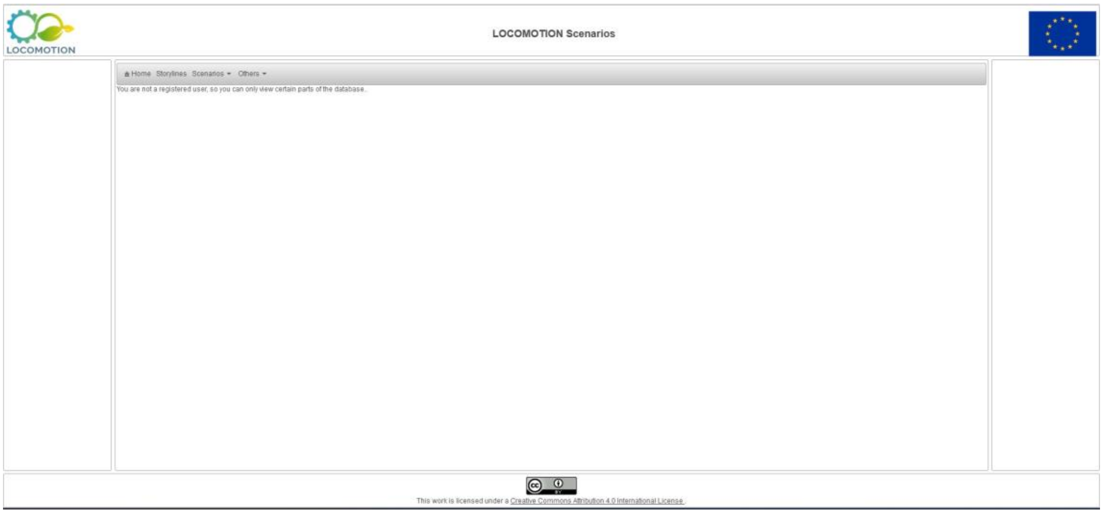


Figure 71: Home page of the WILIAM Database of selected simulated scenarios and results.

Before accessing the features of the Storylines and Scenarios, the user can access the section "Others" in order to obtain information on the main aspects related to the development of the features of the "Storylines" and "Scenarios" sections.

4.13.2.3.1. OTHERS

The user can **view the "Others" Section** (Figure 72) through:

Menu/ Others

This Section allows the user to view the list of:

- Definitions
- Storyline Overall Goals
- Policy measures
- Hypotheses

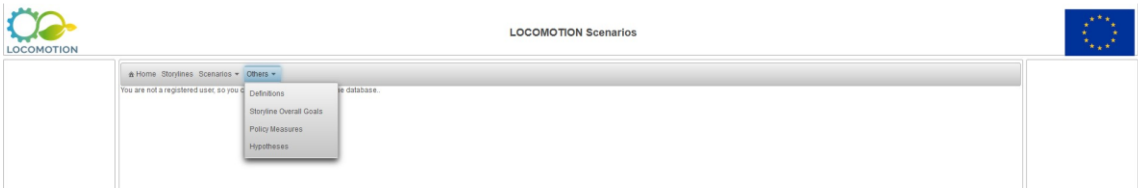


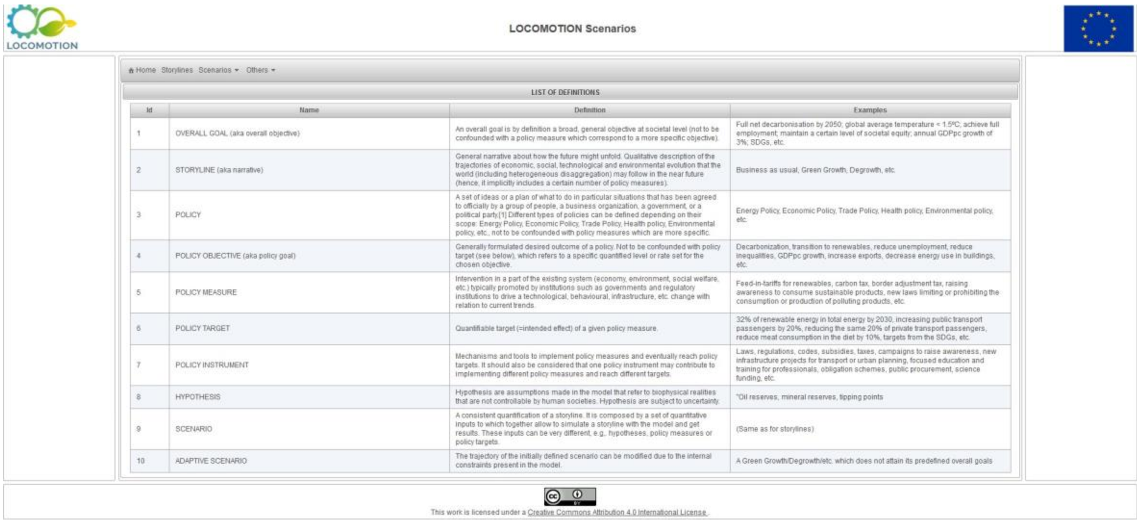
Figure 72: Others List main page.

DEFINITIONS

The user can **access the Definitions** list (Figure 73) through:

Menu/ Others/ Definitions

In this page the user can view the list of common Definitions in LOCOMOTION with examples.



LOCOMOTION Scenarios

Home Storylines Scenarios Others

LIST OF DEFINITIONS

Id	Name	Definition	Examples
1	OVERALL GOAL (aka overall objective)	An overall goal is by definition a broad, general objective at societal level (not to be confounded with a policy measure which correspond to a more specific objective).	Full net decarbonisation by 2050; global average temperature < 1.5ºC; achieve full employment; maintain a certain level of societal equity; annual GDPpc growth of 3%; SDGs, etc.
2	STORYLINE (aka narrative)	General narrative about how the future might unfold. Qualitative description of the trajectories of economic, social, technological and environmental evolution that the world (including heterogeneous disaggregation) may follow in the near future (hence, it implicitly includes a certain number of policy measures).	Business as usual, Green Growth, Degrowth, etc
3	POLICY	A set of ideas or a plan of what to do in particular situations that has been agreed to officially by a group of people, a business organization, a government, or a political party.[1] Different types of policies can be defined depending on their scope: Energy Policy, Economic Policy, Trade Policy, Health policy, Environmental policy, etc., not to be confounded with policy measures which are more specific.	Energy Policy, Economic Policy, Trade Policy, Health policy, Environmental policy, etc.
4	POLICY OBJECTIVE (aka policy goal)	Generally formulated desired outcome of a policy. Not to be confounded with policy target (see below), which refers to a specific quantified level or rate set for the chosen objective.	Decarbonization, transition to renewables, reduce unemployment, reduce inequalities, GDPpc growth, increase exports, decrease energy use in buildings, etc.
5	POLICY MEASURE	Intervention in a part of the existing system (economy, environment, social welfare, etc.) typically promoted by institutions such as governments and regulatory institutions to drive a technological, behavioural, infrastructure, etc. change with relation to current trends.	Feed-in-tariffs for renewables, carbon tax, border adjustment tax, raising awareness to consume sustainable products, new laws limiting or prohibiting the consumption or production of polluting products, etc.
6	POLICY TARGET	Quantifiable target (=intended effect) of a given policy measure.	32% of renewable energy in total energy by 2030, increasing public transport passengers by 20%, reducing the same 20% of private transport passengers, reduce meat consumption in the diet by 10%, targets from the SDGs, etc.
7	POLICY INSTRUMENT	Mechanisms and tools to implement policy measures and eventually reach policy targets. It should also be considered that one policy instrument may contribute to implementing different policy measures and reach different targets.	Laws, regulations, codes, subsidies, taxes, campaigns to raise awareness, new infrastructure projects for transport or urban planning, focused education and training for professionals, obligation schemes, public procurement, science funding, etc.
8	HYPOTHESIS	Hypothesis are assumptions made in the model that refer to biophysical realities that are not controllable by human societies. Hypothesis are subject to uncertainty.	"Oil reserves, mineral reserves, tipping points
9	SCENARIO	A consistent quantification of a storyline. It is composed by a set of quantitative inputs to which together allow to simulate a storyline with the model and get results. These inputs can be very different, e.g., hypotheses, policy measures or policy targets.	(Same as for storylines)
10	ADAPTIVE SCENARIO	The trajectory of the initially defined scenario can be modified due to the internal constraints present in the model.	A Green Growth/Degrowth/etc. which does not attain its predefined overall goals

This work is licensed under a Creative Commons Attribution 4.0 International License.

Figure 73: Definitions List main page.

STORYLINE OVERALL GOALS

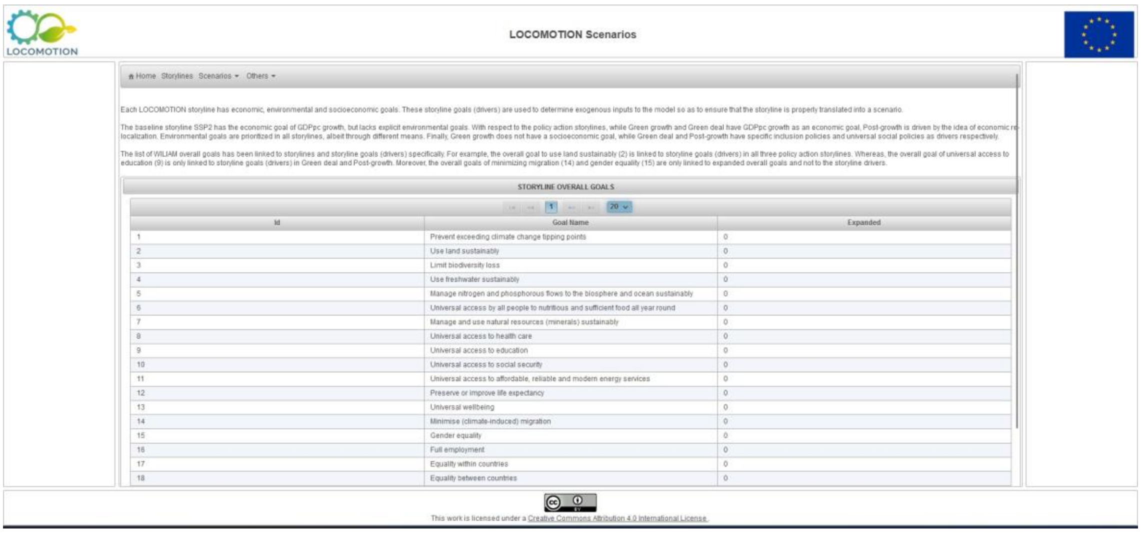
The user can **access the Storyline Overall Goals** list (Figure 74) through:

Menu/ Others/ Storyline Overall Goals

The user can view the existing Storyline Overall Goals. Each LOCOMOTION storyline has economic, environmental and socioeconomic goals. These storyline goals (drivers) are used to determine exogenous inputs to the model so as to ensure that the storyline is properly translated into a scenario.

The baseline storyline SSP2 has the economic goal of GDPpc growth, but lacks explicit environmental goals. With respect to the policy action storylines, while Green growth and Green deal have GDPpc growth as an economic goal, Post-growth is driven by the idea of economic re-localization. Environmental goals are prioritized in all storylines, albeit through different means. Finally, Green growth does not have a socioeconomic goal, while Green deal and Post-growth have specific inclusion policies and universal social policies as drivers respectively.

The list of WILIAM overall goals has been linked to storylines and storyline goals (drivers) specifically. For example, the overall goal to use land sustainably (2) is linked to storyline goals (drivers) in all three policy action storylines. Whereas, the overall goal of universal access to education (9) is only linked to storyline goals (drivers) in Green deal and Post-growth. Moreover, the overall goals of minimizing migration (14) and gender equality (15) are only linked to expanded overall goals and not to the storyline drivers.


LOCOMOTION Scenarios

Home Storylines Scenarios Others

Each LOCOMOTION storyline has economic, environmental and socioeconomic goals. These storyline goals (drivers) are used to determine exogenous inputs to the model so as to ensure that the storyline is properly translated into a scenario.

The baseline storyline SSP2 has the economic goal of GDPpc growth, but lacks explicit environmental goals. With respect to the policy action storylines, while Green growth and Green deal have GDPpc growth as an economic goal, Post-growth is driven by the idea of economic re-localization. Environmental goals are prioritized in all storylines, albeit through different means. Finally, Green growth does not have a socioeconomic goal, while Green deal and Post-growth have specific inclusion policies and universal social policies as drivers respectively.

The list of WILIAM overall goals has been linked to storylines and storyline goals (drivers) specifically. For example, the overall goal to use land sustainably (2) is linked to storyline goals (drivers) in all three policy action storylines. Whereas, the overall goal of universal access to education (9) is only linked to storyline goals (drivers) in Green deal and Post-growth. Moreover, the overall goals of minimizing migration (14) and gender equality (15) are only linked to expanded overall goals and not to the storyline drivers.

STORYLINE OVERALL GOALS

Id	Goal Name	Expanded
1	Prevent exceeding climate change tipping points	0
2	Use land sustainably	0
3	Limit biodiversity loss	0
4	Use freshwater sustainably	0
5	Manage nitrogen and phosphorous flows to the biosphere and ocean sustainably	0
6	Universal access by all people to nutritious and sufficient food all year round	0
7	Manage and use natural resources (minerals) sustainably	0
8	Universal access to health care	0
9	Universal access to education	0
10	Universal access to social security	0
11	Universal access to affordable, reliable and modern energy services	0
12	Preserve or improve life expectancy	0
13	Universal wellbeing	0
14	Minimise (climate-induced) migration	0
15	Gender equality	0
16	Full employment	0
17	Equality within countries	0
18	Equality between countries	0

This work is licensed under a Creative Commons Attribution 4.0 International License.

Figure 74: Storyline Overall Goals List main page.

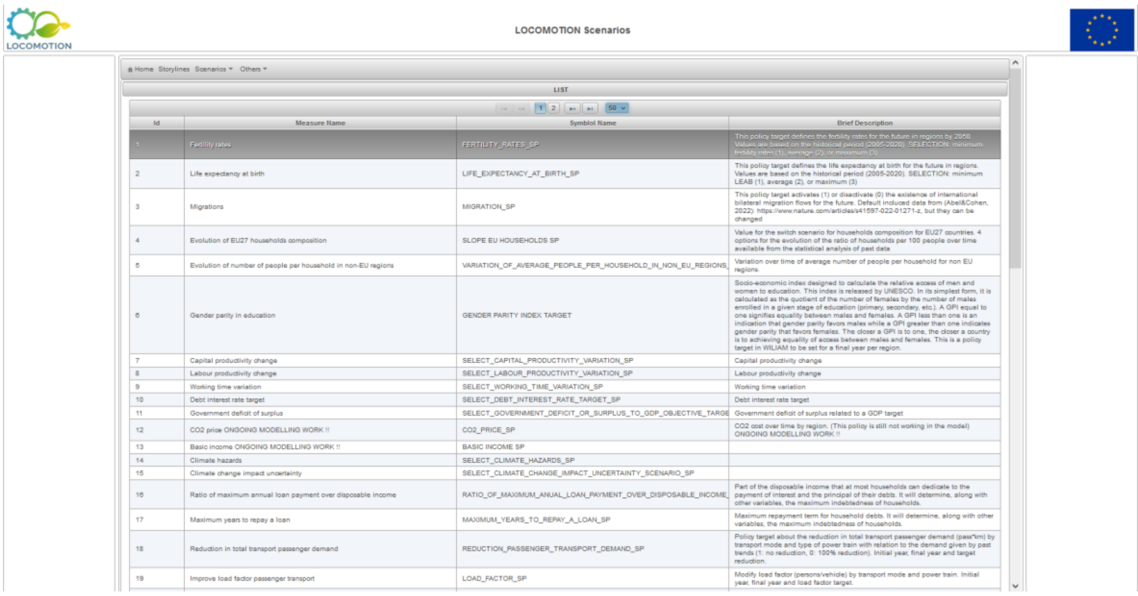
This list aims to include the complete list of Storyline Overall Goals that have been selected in WILIAM. The link between Storyline Overall Goals and each of the Storylines is then performed through:

Menu/ Storylines (for details see section Storylines).

POLICY MEASURES

The user can **access the list of Policy Measures** (Figure 75) through:

Menu/ Others/ Policy Measures


LOCOMOTION Scenarios

Home Storylines Scenarios Others

LIST

Id	Measure Name	Symblol Name	Brief Description
1	Fertility rates	FERTILITY_RATES_SP	This policy target defines the fertility rates for the future in regions by 2050. Values are based on the historical period (2005-2020). SELECTION: minimum fertility rates (1), average (2), or maximum (3)
2	Life expectancy at birth	LIFE_EXPECTANCY_AT_BIRTH_SP	This policy target defines the life expectancy at birth for the future in regions. Values are based on the historical period (2005-2020). SELECTION: minimum LEAB (1), average (2), or maximum (3)
3	Migrations	MIGRATION_SP	This policy target activates (1) or disactivate (0) the existence of international bilateral migration flows for the future. Default incluced data from (Abel&Cohen, 2022): https://www.nature.com/articles/s41597-022-01271-z, but they can be changed
4	Evolution of EU27 households composition	SLOPE EU HOUSEHOLDS SP	Value for the switch scenario for households composition for EU27 countries. 4 options for the evolution of the ratio of households per 100 people over time available from the statistical analysis of past data
5	Evolution of number of people per household in non-EU regions	VARIATION_OF_AVERAGE_PEOPLE_PER_HOUSEHOLD_IN_NON_EU_REGIONS_	Variation over time of average number of people per household for non EU regions.
6	Gender parity in education	GENDER PARITY INDEX TARGET	Socio-economic index designed to calculate the relative access of men and women to education. This index is released by UNESCO. In its simplest form, it is calculated as the quotient of the number of females by the number of males enrolled in a given stage of education (primary, secondary, etc.). A GPI equal to one signifies equality between males and females. A GPI less than one is an indication that gender parity favors males while a GPI greater than one indicates gender parity that favors females. The closer a GPI is to one, the closer a country is to achieving equality of access between males and females. This is a policy target in WILIAM to be set for a final year per region.
7	Capital productivity change	SELECT_CAPITAL_PRODUCTIVITY_VARIATION_SP	Capital productivity change
8	Labour productivity change	SELECT_LABOUR_PRODUCTIVITY_VARIATION_SP	Labour productivity change
9	Working time variation	SELECT_WORKING_TIME_VARIATION_SP	Working time variation
10	Debt interest rate target	SELECT_DEBT_INTEREST_RATE_TARGET_SP	Debt interest rate target
11	Government deficit of surplus	SELECT_GOVERNMENT_DEFICIT_OR_SURPLUS_TO_GDP_OBJECTIVE_TARGE	Government deficit of surplus related to a GDP target
12	CO2 price ONGOING MODELLING WORK !!	CO2_PRICE_SP	CO2 cost over time by region. (This policy is still not working in the model) ONGOING MODELLING WORK !!
13	Basic income ONGOING MODELLING WORK !!	BASIC INCOME SP	
14	Climate hazards	SELECT_CLIMATE_HAZARDS_SP	
15	Climate change impact uncertainty	SELECT_CLIMATE_CHANGE_IMPACT_UNCERTAINTY_SCENARIO_SP	
16	Ratio of maximum annual loan payment over disposable income	RATIO_OF_MAXIMUM_ANUAL_LOAN_PAYMENT_OVER_DISPOSABLE_INCOME_	Part of the disposable income that at most households can dedicate to the payment of interest and the principal of their debts. It will determine, along with other variables, the maximum indebtedness of households.
17	Maximum years to repay a loan	MAXIMUM_YEARS_TO_REPAY_A_LOAN_SP	Maximum repayment term for household debts. It will determine, along with other variables, the maximum indebtedness of households.
18	Reduction in total transport passenger demand	REDUCTION_PASSENGER_TRANSPORT_DEMAND_SP	Policy target about the reduction in total transport passenger demand (pass*km) by transport mode and type of power train with relation to the demand given by past trends (1: no reduction, 0: 100% reduction). Initial year, final year and target reduction.
19	Improve load factor passenger transport	LOAD_FACTOR_SP	Modify load factor (persons/vehicle) by transport mode and power train. Initial year, final year and load factor target.

Figure 75: Policy Measures List main page.

This page provides the user with an overview of the list of policy measures featured in WILIAM, information on the Symbol associated with each policy measure and a brief narrative description of the policy measure.

HYPOTHESES

The user can **access the list of Hypotheses** (Figure 76) through:

Menu/ Others/ Hypotheses

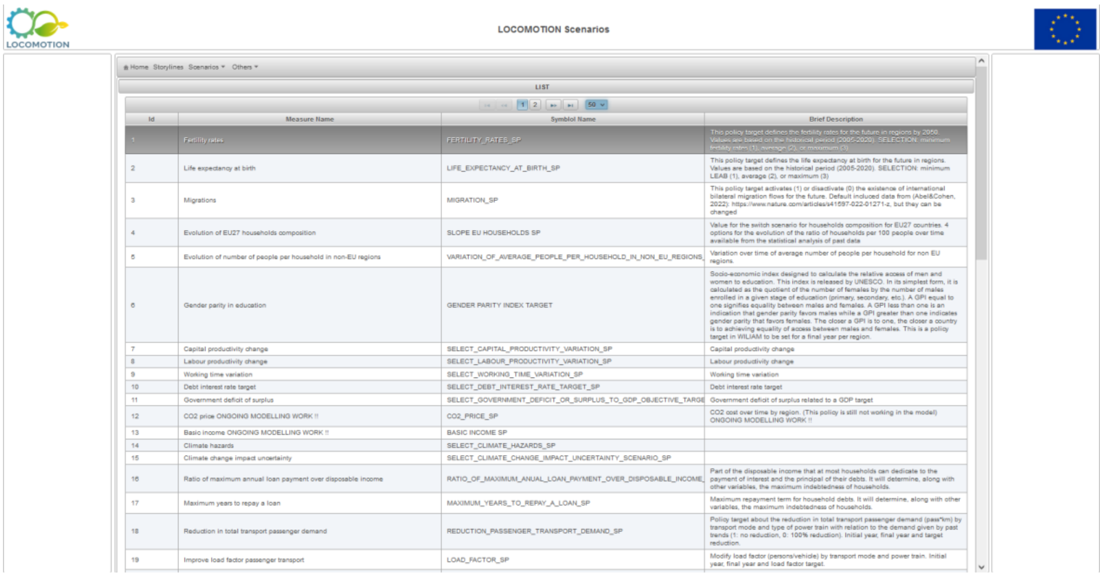


Figure 76: Hypotheses List main page.

This page provides the user with an overview of the list of hypotheses featured in WILIAM, information on the Symbol associated with each hypothesis and a brief narrative description of the hypothesis.

4.13.2.3.2. STORYLINES

View existing Storylines

The user can **view existing Storylines** (Figure 77) through:

Menu/ Storylines

Figure 77: Storylines List main page.

The user can **view all the details of a Storyline** through:

Menu/ Storyline: Select an Storyline in list and press "View details"

Storyline Fields

The user can **view the fields of a Storyline**, as it can be seen in Figure 78. This includes the name of the storyline, as well as a narrative description of the storyline.

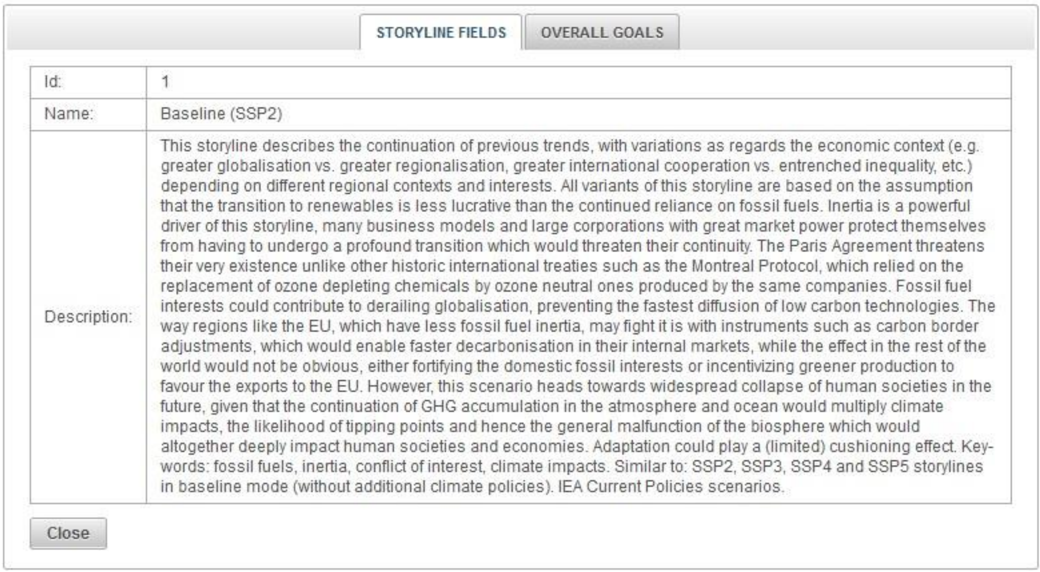


Figure 78: Storyline Fields page.

Storyline Overall Goals

The user can **view the overall goals linked to a Storyline**, as it can be seen in Figure 79.

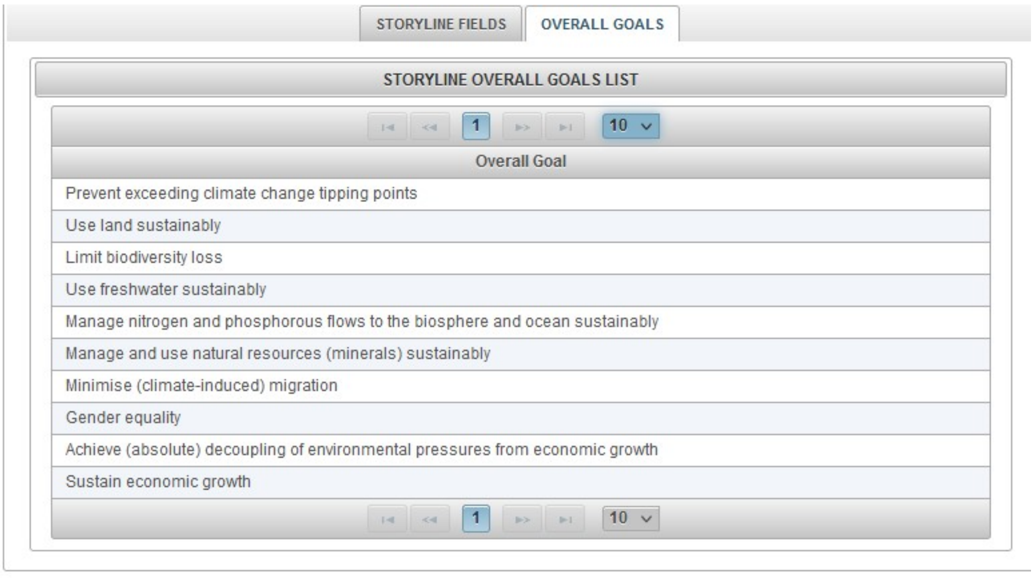


Figure 79: Storyline Overall Goals page.

4.13.2.3.3. SCENARIOS

SCENARIOS PARAMETERS

The user can **view the Scenarios and the Scenario input parameters** (Figure 80) through:

Menu/ Scenarios/ Parameters

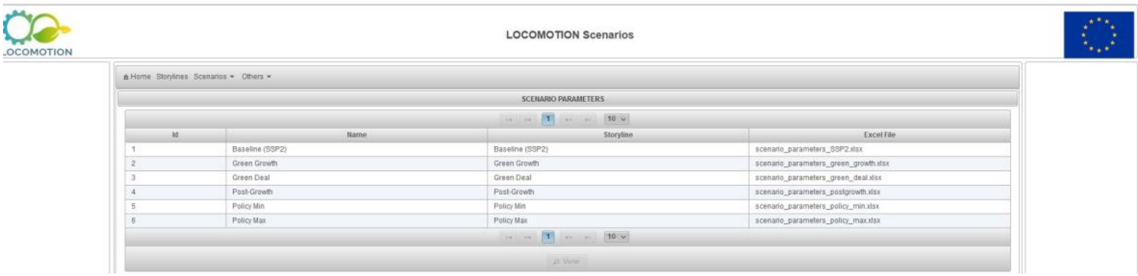


Figure 80: Scenarios List main page.

In this page the user can view the list of scenarios. The user can **view the details of a Scenario** (Figure 81) through:

Menu/ Scenarios/ Parameters: Select a Scenario and in the bottom of the page press "View"

In this window, the user has several options, which are presented below.

Scenario fields

The user can **view the Scenario fields**, i.e., Scenario Name, the narrative Description of the Scenario and the Storyline that is associated with the Scenario.

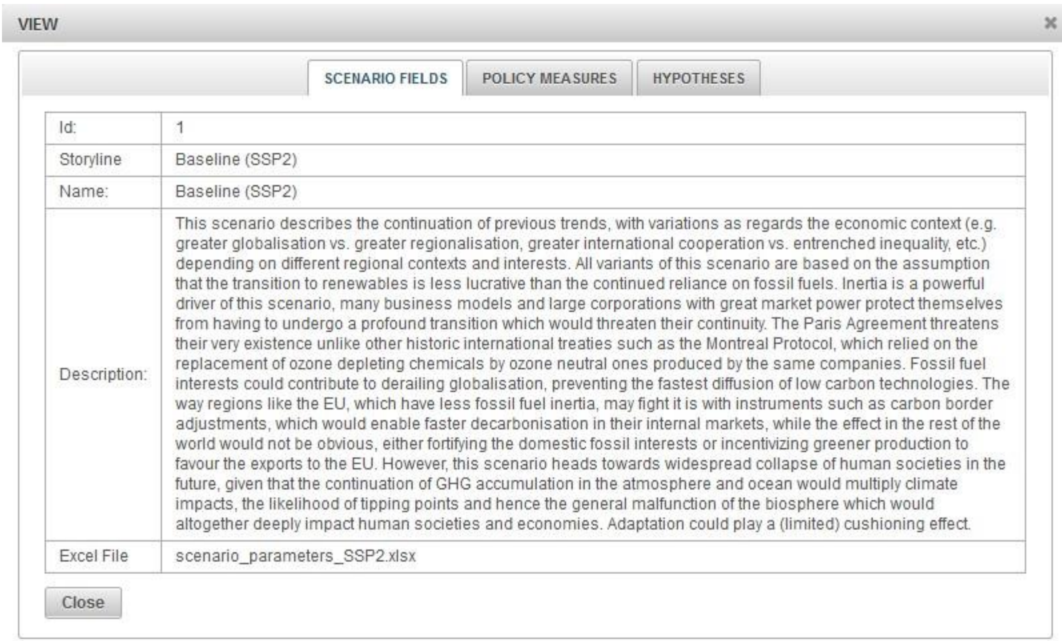
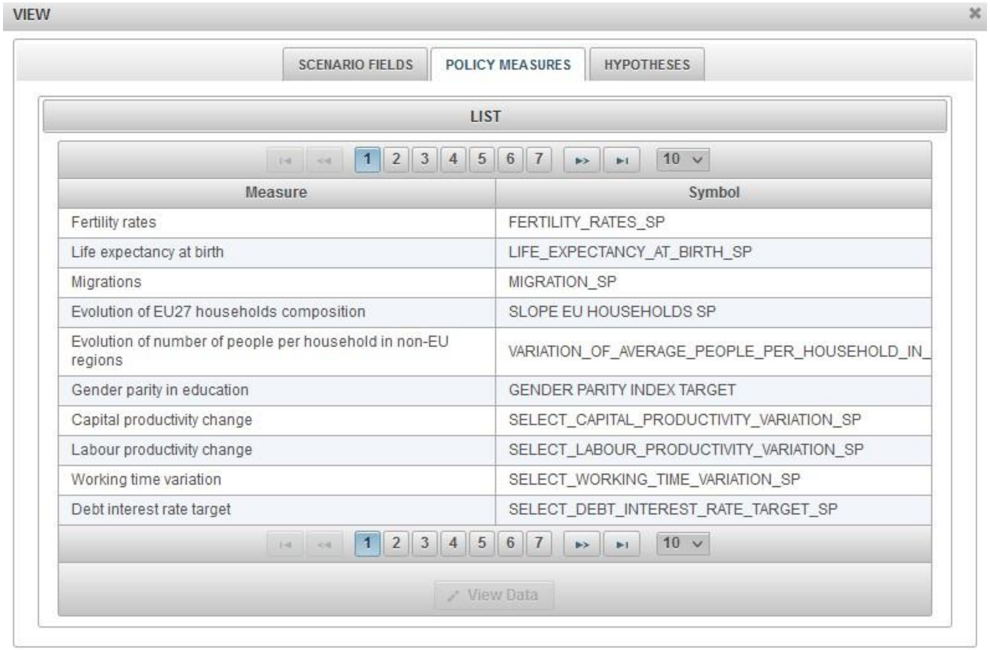


Figure 81: View Scenario page.

Policy measures

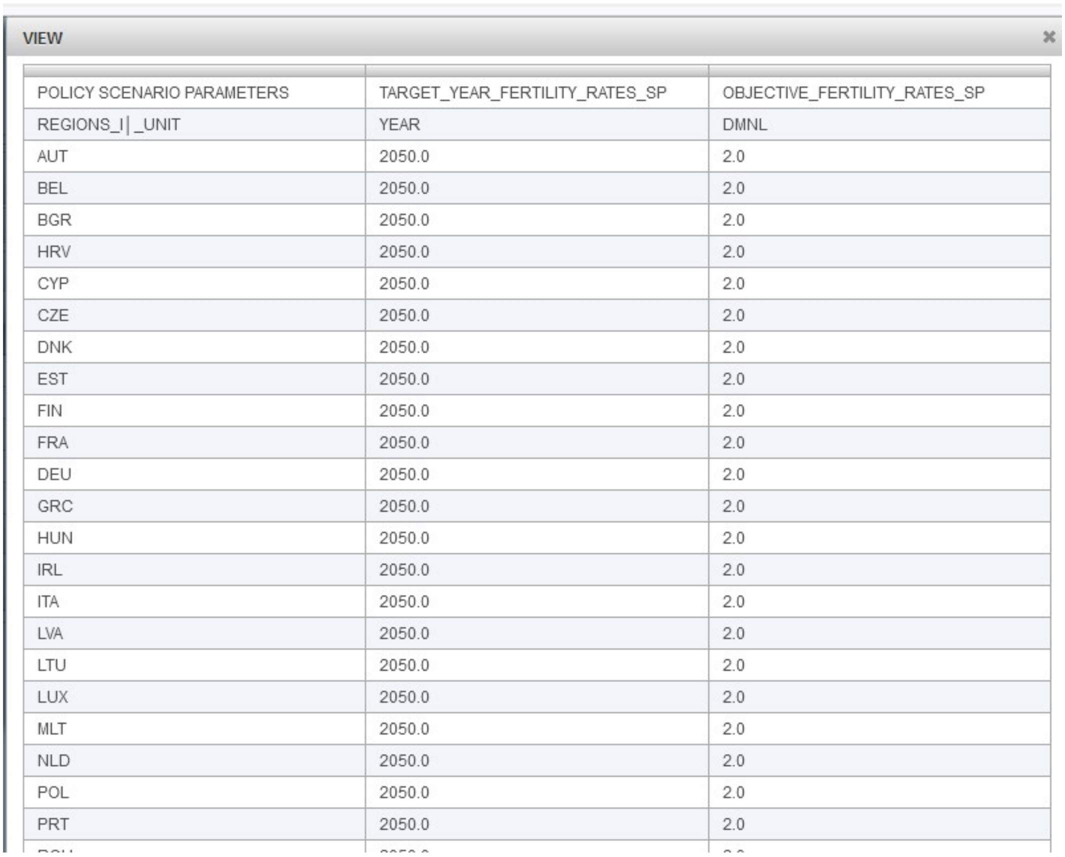
The user can **view the list of Policy measures** and the Scenario input values for each policy measure (Figure 82).



Measure	Symbol
Fertility rates	FERTILITY_RATES_SP
Life expectancy at birth	LIFE_EXPECTANCY_AT_BIRTH_SP
Migrations	MIGRATION_SP
Evolution of EU27 households composition	SLOPE EU HOUSEHOLDS SP
Evolution of number of people per household in non-EU regions	VARIATION_OF_AVERAGE_PEOPLE_PER_HOUSEHOLD_IN_
Gender parity in education	GENDER PARITY INDEX TARGET
Capital productivity change	SELECT_CAPITAL_PRODUCTIVITY_VARIATION_SP
Labour productivity change	SELECT_LABOUR_PRODUCTIVITY_VARIATION_SP
Working time variation	SELECT_WORKING_TIME_VARIATION_SP
Debt interest rate target	SELECT_DEBT_INTEREST_RATE_TARGET_SP

Figure 82: View Policy Measures associated with a Scenario page.

The user can **view the input parameters of a Scenario** (Figure 83) through the Policy Measures page: Select a Policy Measure and in the bottom of the page press “View Data”.

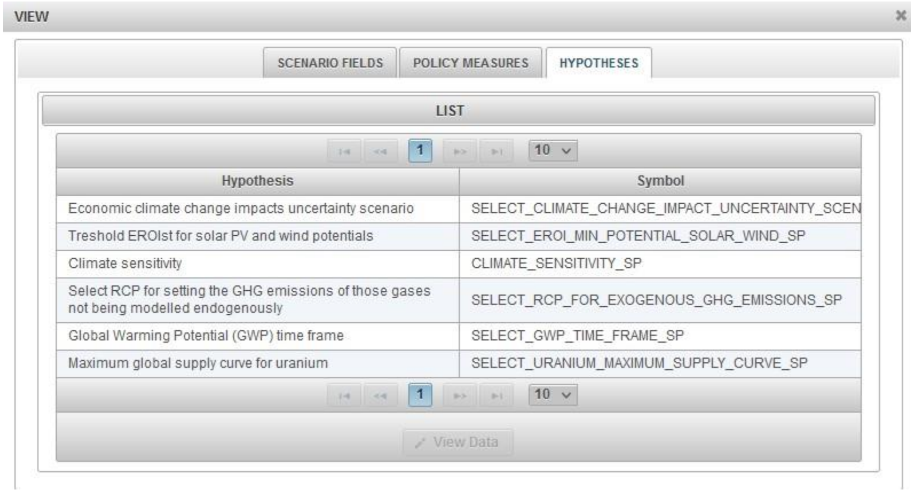


POLICY SCENARIO PARAMETERS	TARGET_YEAR_FERTILITY_RATES_SP	OBJECTIVE_FERTILITY_RATES_SP
REGIONS_I| _UNIT	YEAR	DMNL
AUT	2050.0	2.0
BEL	2050.0	2.0
BGR	2050.0	2.0
HRV	2050.0	2.0
CYP	2050.0	2.0
CZE	2050.0	2.0
DNK	2050.0	2.0
EST	2050.0	2.0
FIN	2050.0	2.0
FRA	2050.0	2.0
DEU	2050.0	2.0
GRC	2050.0	2.0
HUN	2050.0	2.0
IRL	2050.0	2.0
ITA	2050.0	2.0
LVA	2050.0	2.0
LTU	2050.0	2.0
LUX	2050.0	2.0
MLT	2050.0	2.0
NLD	2050.0	2.0
POL	2050.0	2.0
PRT	2050.0	2.0

Figure 83: Example of Policy Measure inputs for a Scenario view.

Scenario Hypotheses

The user can **view the list of Scenario Hypotheses** and the Scenario Hypotheses input values for each hypothesis (Figure 84).



Hypothesis	Symbol
Economic climate change impacts uncertainty scenario	SELECT_CLIMATE_CHANGE_IMPACT_UNCERTAINTY_SCEN
Treshold EROIst for solar PV and wind potentials	SELECT_EROI_MIN_POTENTIAL_SOLAR_WIND_SP
Climate sensitivity	CLIMATE_SENSITIVITY_SP
Select RCP for setting the GHG emissions of those gases not being modelled endogenously	SELECT_RCP_FOR_EXOGENOUS_GHG_EMISSIONS_SP
Global Warming Potential (GWP) time frame	SELECT_GWP_TIME_FRAME_SP
Maximum global supply curve for uranium	SELECT_URANIUM_MAXIMUM_SUPPLY_CURVE_SP

Figure 84: View Scenario Hypotheses page.

The user can **view the input parameters of a Scenario Hypothesis** (Figure 85) through the Hypotheses page: Select a Hypothesis and in the bottom of the page press “View Data”.

_VALUE_OPTIONS | Dmnl	_OPTIONS
0.0	Maximum damages (the statistic measure 'maximum' is used to calibrate the damage function)
1.0	Minimum damages (the statistic measure 'minimum' is used to calibrate the damage function)
2.0	Median damages (the statistic measure 'median' is used to calibrate the damage function)
3.0	Average damages (the statistic measure 'mean' is used to calibrate the damage function)
_VALUE_SELECTED_OPTION	2.0

Figure 85: Example of Hypotheses inputs for a Scenario view.

RESULTS

The user can **view the Scenario Results** (Figure 86) through:

Menu/ Scenarios/ Results

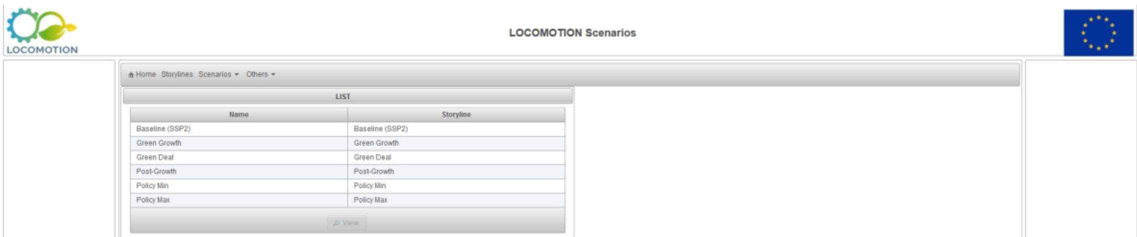


Figure 86: Scenarios Results main page.

In this page the user can view the list of scenarios. The user can **view the Results of a Scenario** through:

Menu/ Scenarios/ Results: Select a Scenario and in the bottom of the page press “View”

A new table appears with the available model simulations, the date when they were obtained and information on the model version (Figure 87).

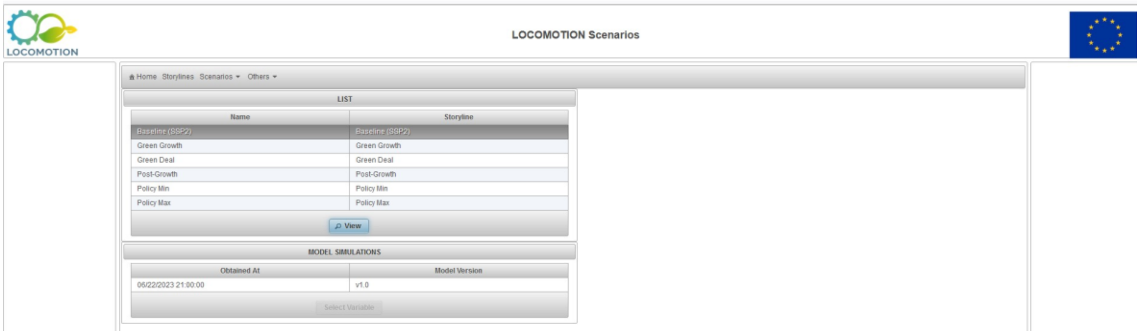


Figure 87: Scenarios Results selection page.

The user can select the model simulation available and press "Select Variable". A new window appears, as it can be seen in Figure 88. In this window the user has the option to use search criteria to filter a Variable. These criteria include the Variable name and the main module of the Variable. The search can be flexible or not, i.e., the Variable name must match exactly or partially the search text. Otherwise, the complete list of Variables appears.

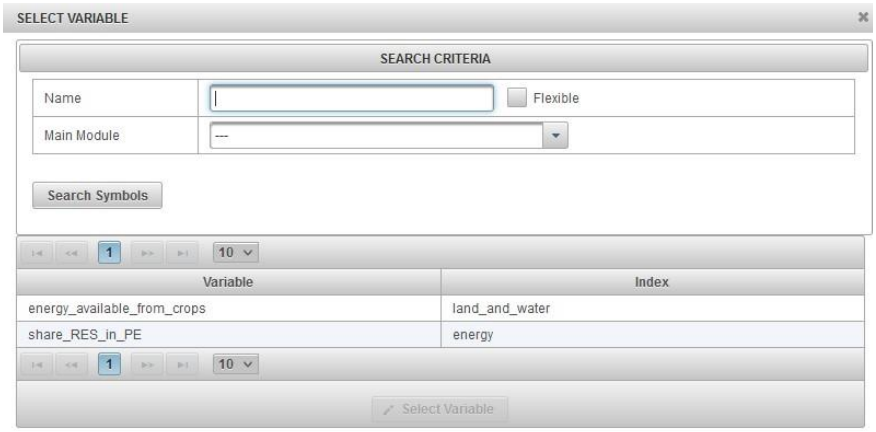


Figure 88: Scenario Results Variable selection page.

The user can select a Variable out of the list and press "Select Variable". A new table appears which provides information on the Variable name, Unit and associated Module (Figure 89). The user can press "View Data".

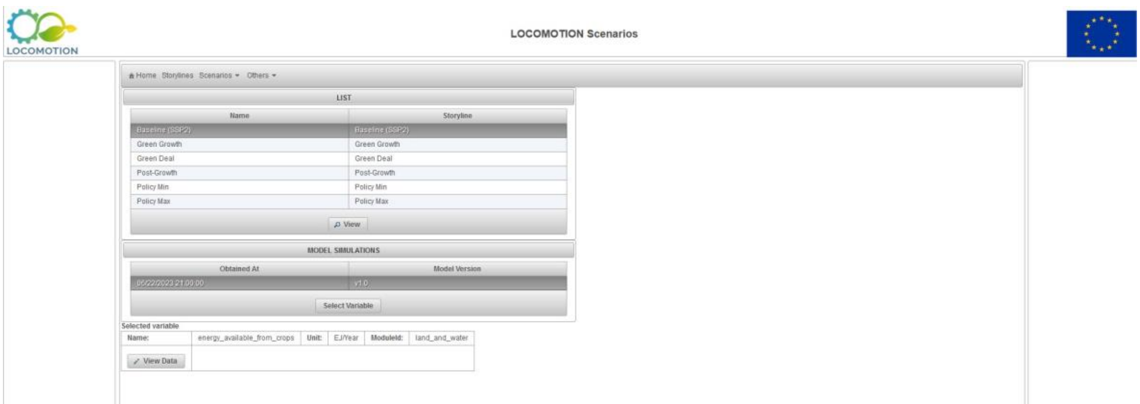


Figure 89: Scenario Results Variable page.

The user can view the data associated with the selected Variable as a time series Table (Figure 90), as a Line Graph (Figure 91) or as a Graph with data for 2050 (Figure 92).

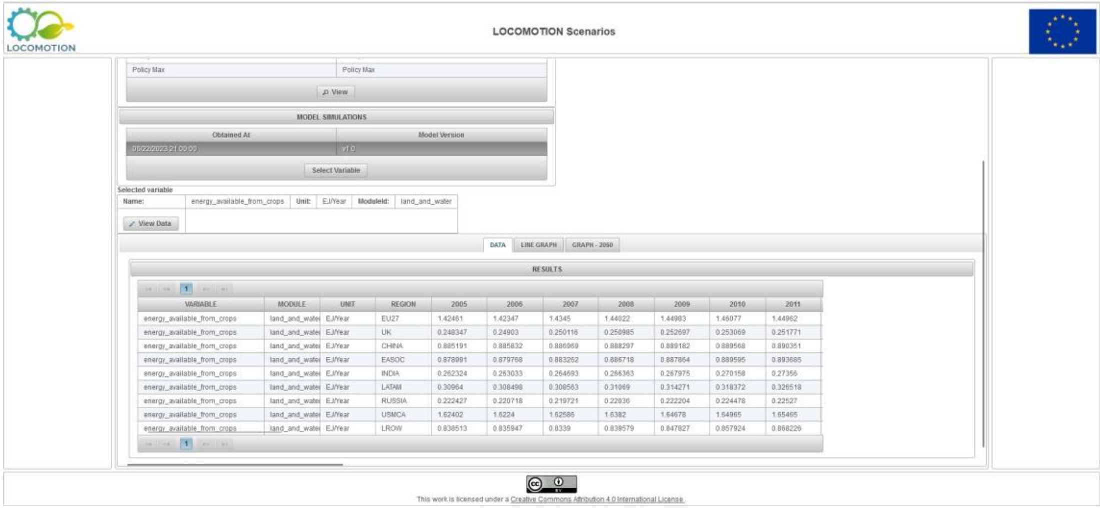


Figure 90: Scenario Results Table.

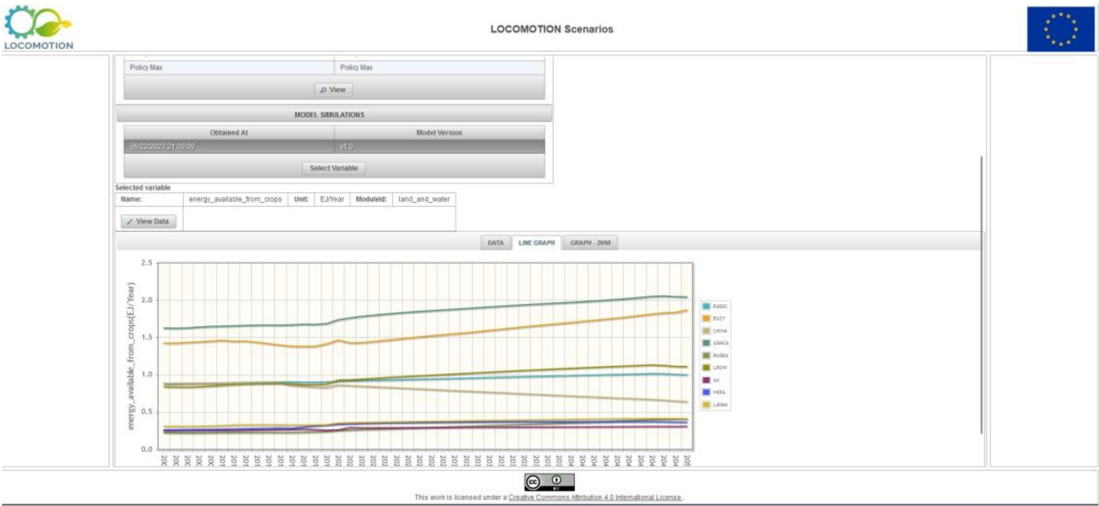


Figure 91: Scenario Results Line Graph.

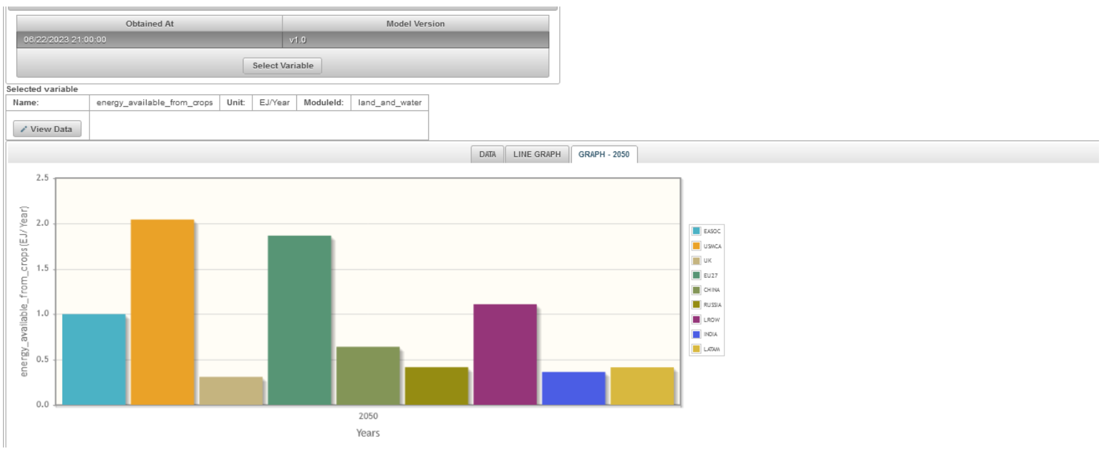


Figure 92: Scenario Results Graph 2050.

4.13.3. FULFILLMENT OF DATA REQUIREMENTS FROM COMMISSION

Data quality does not only foresee that data are fit for purpose, but also that data needs to be Findable, Accessible, Interoperable and Reusable (FAIR), so that whatever is obtained from it is totally transparent and can be replicated and traced back to its origin. To this regard, good quality and rich metadata plays an essential role. The FAIR principles (Wilkinson, 2016) were developed to guide all individuals involved in the data lifecycle in order to make data findable, accessible, interoperable and reusable both for humans and machines.

In order to ensure that the data and metadata generated for the WILIAM model are of high quality and as FAIR as possible a set of measures has been implemented to enforce FAIRness of the project (meta)data.

4.13.3.1. FINDABLE

Findability of the project (meta)data is achieved through the registration of the (meta)data resources in a data repository. Both the WILIAM model (Vensim® and Python versions) and the database h will be uploaded in https://zenodo.org and will obtain a unique identifier. The shared database and open database management system, both the Data Dictionary and the WILIAM Database of selected simulated scenarios and results, have been uploaded in zenodo (https://zenodo.org/record/8214171) and they have received doi: 10.5281/zenodo.8214171. The WILIAM model (Vensim® and Python versions) will also be uploaded in zenodo.

In addition, datasets are described with 38 metadata fields, of which at least 3 assist in locating the data and the sources, i.e., “Values read from”, which indicates the file name, tab name, and cell names where the data is located inside the Inputs spreadsheet file, the “Source data” indicates the data sources from which the dataset was generated, and the “Link”, provides a URL to the original data source from which the source data was obtained.

4.13.3.2. ACCESSIBLE

The accessibility of the project (meta)data is achieved through making the WILIAM model, the Input files and the Data Dictionary and the Database of selected simulated scenarios and results available through the project website without specialised or proprietary tools, as apart from the Vensim® version of the model, there is also a Python version of the model, which is available as open-source. The protocol is open and free in order to maximise data reuse and retrieval. In detail, the WILIAM model in Vensim® and Python (pywiliam) is available through the project website with URL: https://www.locomotion-h2020.eu/locomotion-models/locomotion-iams/. The code is available with URL: https://gitlab-locomotion.infor.uva.es/locomotion/wiliam and https://gitlab-locomotion.infor.uva.es/locomotion/pywiliam. The Input files of the model, which include the data, are saved in xlsx format, which is readable using free and open-source spreadsheet software (e.g., LibreOffice). Manual human intervention is required to access the data. Metadata is accessible and retrievable through the Data Dictionary using the standard http(s) communications protocol, where users can make requests by Symbol name or by list of Symbol names. When the Symbol name is left blank, all Symbols are returned. The Data Dictionary (webclient and API) allows retrieving the metadata through a spreadsheet, CSV and xml open source and free format. Access to the (meta)data is guaranteed by making them available through the zenodo repository, available through https://zenodo.org/record/8214171.

4.13.3.3. INTEROPERABLE

The interoperability for the data and metadata developed in the LOCOMOTION project is ensured by (meta)data that make use of commonly used controlled vocabularies, ontologies and thesauri and adopt the most commonly used international nomenclature. Information on acronyms, adjectives and semantic rules are available through the Data Dictionary that has been developed in the project, which is currently hosted in CRES with URL: http://www.cres.gr/LOCOMOTION_client, and it is available through the LOCOMOTION project website with URL: https://www.locomotion-h2020.eu/locomotion-models/database/. Also, through the Data Dictionary, the relations between Symbols and datasets are specified and all datasets are cited.

4.13.3.4. REUSABLE

The LOCOMOTION project adopts an open-source/open-access framework for the WILIAM model and (meta)data in order to ensure the reusability of data in the project and provides rich metadata to describe the context of data generation, as detailed in section 4.13.1.2.3 Metadata. The WILIAM model is delivered in a Vensim® version and a Python version, which is a widely used, open-source programming language.

The WILIAM model (Vensim® version), as well as the WILIAM model in Python (pywiliam), the Model Explorer, the Model Analyzer and the CROSSROADS game are available in the project website under an MIT license and free of charge. The MIT License is a permissive license which grants permission "free of charge to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions: The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software. THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE" (Open source initiative, n.d.).

The Data Dictionary and the Database of selected simulated scenarios and results are available under a CC BY license. This license lets others distribute, remix, adapt, and build upon your work, even commercially, as long as they credit you for the original creation. This is the most accommodating of licenses offered and it is recommended for maximum dissemination and use of licensed materials (Creative Commons, n.d.).

In order to ensure reusability both the WILIAM model and the database willl be uploaded in https://zenodo.org and will obtain a unique identifier. The shared database and open database management system, both the Data Dictionary and the WILIAM Database of selected simulated scenarios and results, have been uploaded in zenodo (https://zenodo.org/record/8214171) and they have received doi: 10.5281/zenodo.8214171. The WILIAM model (Vensim® and Python versions) will also be uploaded in zenodo.

4.13.4. (META)DATA QUALITY CHECKS

The data and metadata quality checks are performed using the spreadsheet files used as Input to the models and the Data Dictionary, respectively. The quality goals for each quality dimension that are pursued through the data quality assessment include a system to classify the identified issues according to their severity (WARNING, ERROR, CRITICAL).

(Meta)data quality checks have been automatised with Python code, so that they can be executed several times until the end of the project. A dedicated open-source Python library (SDQC) has been developed, which automatises the quality checks on the data stored in the model Input files. On the other hand, the

Great Expectations library (https://docs.greatexpectations.io/docs/) has been used to parametrise and automatise the metadata quality checks, reading from the Data Dictionary.

The code of the SDQC library can be found in this repository: https://gitlab.com/eneko.martin.martinez/sdqc and is installable from the Python Package Index (https://pypi.org/project/sdqc/) using the package installer for Python (pip). A new menu is available on the Data Dictionary where the results of the metadata quality checks run twice per month may be consulted (http://www.cres.gr/LOCOMOTION_client/faces/qualityCheck.xhtml?jftfdi=&jffi=%2FqualityCheck.xhtml).

A brief overview of the performed data quality checks is presented in Table 11.

Table 11: Types of data quality checks.

Check type	Parameters of the check
Missing values in data	Any missing values in the data
Missing values in series	Any missing values in the series
Outliers	3 times the standard deviation of the whole dataset (including all dimensions)
Series increment type	All series data should follow a linear trend
Series monotony	All series data should increase monotonically (years numbers should increase in series data)
Series range	Series values should range between 2005 and 2050, which is the time-span simulated by WILIAM

The criteria to classify the issues' severity is shown in Table 12.

Table 12: Criteria to classify data quality issues by severity.

Check type	Error type
Missing values in data	WARNING: <= 20% of issues of this type ERROR: 20% < % issues of this type <= 50 % CRITICAL: >50% issues of this type:
Missing values in series	
Outliers	
Series increment type	WARNING
Series monotony	CRITICAL
Series range	ERROR

The metadata quality checks that run in the Data Dictionary by level of severity are presented in Table 13. The checks focus on three levels, metadata, Symbols and Indices.

Table 13: Metadata quality checks in Data Dictionary by level of severity.

	Metadata CRITICAL	
	None	
	Metadata ERROR	
1	license	1) values must never be null
2	license_code	1) values must never be null
3	license_code _detailed	1) values must belong to this set: CC BY 1.0 CC BY 2.0 CC BY 2.5 CC BY 3.0 CC BY 4.0 CC BY-SA 1.0 CC BY-SA 2.0 CC BY-SA 2.5 CC BY-SA 3.0 CC BY-SA 4.0 CC BY-NC 1.0 CC BY-NC 2.0 CC BY-NC 2.5 CC BY-NC 3.0 CC BY-NC 4.0 CC BY-NC-SA 1.0 CC

		BY-NC-SA 2.0 CC BY-NC-SA 2.5 CC BY-NC-SA 3.0 CC BY-NC-SA 4.0 CC BY-ND 1.0 CC BY-ND 2.0 CC BY-ND 2.5 CC BY-ND 3.0 CC BY-ND 4.0 CC BY-NC-ND 1.0 CC BY-NC-ND 2.0 CC BY-NC-ND 2.5 CC BY-NC-ND 3.0 CC BY-NC-ND 4.0 CC0 CC0 1.0 CC BY-NC-SA 3.0 IGO CDLA-Permissive-1.0 CDLA-Permissive-2.0 CDLA-Sharing-1.0 ODC-BY ODC-BY-1.0 ODC-ODbL ODC-ODbL-1.0 PDDL PDDL-1.0Public Domain Other	
4	methodology	1) values must never be null	
5	source	1) values must not be null, at least 100 % of the time	
6	unit	1) values must never be null	
	Metadata WARNING		
1	contact	1) values must never be null	
2	license_code	1) values must not match this regular expression: ^Other$, at least 80 % of the time	
3	reliability	1) values must not be null, at least 70 % of the time	2) values must be greater than or equal to 0 and less than or equal to 100
	Symbol CRITICAL		
1	name	1) must never be null	2) must be unique
	Symbol ERROR		
1	is_validated	1) values must be equal to 1, at least 90 % of the time	
2	name	1) values must match this regular expression: (?!\d)[A-Za-z0-9_]+$	
3	type_in_programming	1) values must never be null	
	Symbol WARNING		
1	definition	1) must not be null at least 90% of the time	2) must be unique
2	excel_info	1) must never be null	2) must be unique
3	project_value_type_id	1) values must not be null, at least 90 % of the time	
4	unit	1) values must not be null, at least 100 % of the time	
	Index definition ERROR		
1	definition	1) must be unique	2) values must never be null
2	display_name	1) must never be null	2) must be unique
3	index_name	1) values must never be null	2) values must match this regular expression: (?!\d)[A-Za-z0-9_]+$
	Index definition WARNING		
1	definition	1) value types must belong to this set: CHAR NCHAR VARCHAR NVARCHAR TEXT NTEXT STRING StringType string str object dtype('O')	

2	display_name	1) value types must belong to this set: CHAR NCHAR VARCHAR NVARCHAR TEXT NTEXT STRING StringType string str object dtype('O')	
3	index_name	1) value types must belong to this set: CHAR NCHAR VARCHAR NVARCHAR TEXT NTEXT STRING StringType string str object dtype('O')	
4	is_validated	1) values must belong to this set: 0 1	2) values must be equal to 1
	Index value ERROR		
1	value_name	1) values must match this regular expression: (?!\d)[A-Za-z0-9]+$	2) values must never be null

4.14. MODEL AVAILABILITY

All the LOCOMOTION project developments that are included in this deliverable will be available through the project website and more specifically through the online space provided on it for each exploitable result of the project:

- WILIAM model: https://www.locomotion-h2020.eu/locomotion-models/locomotion-iams/
- Data Dictionary and results database: https://www.locomotion-h2020.eu/locomotion-models/database/
- Model Analyser: https://www.locomotion-h2020.eu/locomotion-models/model-analyser/
- Global Crossroads II Game: https://www.locomotion-h2020.eu/locomotion-models/global-sustainability-crossroads-ii/
- Model Explorer: https://www.locomotion-h2020.eu/locomotion-models/model-explorer/

In addition, Gitlab, github and Zenodo repositorios will be available to share the code developments of the projects. Considering that some of the results are under development, this repositories will be publicly available in the last month of the project. This strategy will improve the transparency of the project and promotes code sharing between researchers and modelling groups to capitalize LOCOMOTION developments.

CONCLUSIONS

In this document we describe the work done to translate the WILIAM model from Vensim® to Python and to write the WILIAM e-handbook.

For the translation of WILIAM to Python, most efforts were put into improving a pre-existing Python library, named PySD, which automatically translates models written in Vensim® or Stella to Python. Older versions of this library were not production ready, lacked many Vensim® functions and produced difficult to read Python code. In addition, execution performance was substantially slower than in Vensim®. All these problems have been addressed, and the WILIAM model has successfully been translated to Python. In addition, the pywiliam interface has been developed to facilitate launching simulations and plotting results from the pywiliam model. All code produced in this work is open-source and under the MIT license.

The WILIAM e-handbook will serve as a relatively short description of the WILIAM model, which will be shared in workshops and similar events, to provide a quick glance of the model. The writing of the e-handbook was a shared effort by most LOCOMOTION partners. This task started by defining the contents that the e-handbook should have. Then, in order to have a similar structure for all chapters describing

WILIAM modules, a template was made for the contents of a chapter and was sent to all contributing authors. After receiving the authors contributions, this task reviewed the contents and generated a final version of the WILIAM e-handbook. Throughout this document, there is a functional and simplified documentation of the model that helps to understand the modules and approaches used in the model.

REFERENCES

Abel, G. J., & Cohen, J. E. (2022). Bilateral international migration flow estimates updated and refined by sex. *Scientific Data*, *9*(1), 173. https://doi.org/10.1038/s41597-022-01271-z

Anderson, K. (2007). Chapter 12 - Evidence and memory: records services and archives. In S. B. T.-L. in the T.-F. C. Ferguson (Ed.), *Topics in Australasian Library and Information Studies* (pp. 203–219). Chandos Publishing. https://doi.org/https://doi.org/10.1016/B978-1-876938-43-7.50012-0

Anthoff, D., & Tol, R. S. J. (2014). *The Climate Framework For Uncertainty, Negotiation And Distribution (Fund), Technical Description, Version 3.9*.

Auerbach, A. J., & Gorodnichenko, Y. (2012). Measuring the Output Responses to Fiscal Policy. *American Economic Journal: Economic Policy*, *4*(2), 1–27. https://doi.org/10.1257/pol.4.2.1

Barlas, Y. (1996). Formal aspects of model validity and validation in system dynamics. *System Dynamics Review*, *12*(3), 183–210. https://doi.org/https://doi.org/10.1002/(SICI)1099-1727(199623)12:3<183::AID-SDR103>3.0.CO;2-4

Barnes, J., Dove, M., Lahsen, M., Mathews, A., McElwee, P., McIntosh, R., Moore, F., O'Reilly, J., Orlove, B., Puri, R., Weiss, H., & Yager, K. (2013). Contribution of anthropology to the study of climate change. *Nature Climate Change*, *3*(6), 541–544. https://doi.org/10.1038/nclimate1775

Bednar, J., Obersteiner, M., & Wagner, F. (2019). On the financial viability of negative emissions. *Nature Communications*, *10*(1), 1783. https://doi.org/10.1038/s41467-019-09782-x

Blanford, G. J., Kriegler, E., & Tavoni, M. (2014). Harmonization vs. fragmentation: overview of climate policy scenarios in EMF27. *Climatic Change*, *123*(3), 383–396. https://doi.org/10.1007/s10584-013-0951-9

Böck, E., Eggler, L., Rohrer, M., Papagianni, S., Christodoulaki, R., & Taxeri, E. (2020). *LOCOMOTION Deliverabe 8.1: Review of policy opt ions to drive societies towards sustainability*. https://www.LOCOMOTION-h2020.eu/.

Böhringer, C., García-Muros, X., & González-Eguino, M. (2019). A Progressive Environmental Tax Reform for Spain. *Economics of Energy & Environmental Policy*, *8*(2), 141–160. https://www.jstor.org/stable/26780610

Bosetti, V., Carraro, C., Galeotti, M., Massetti, E., & Tavoni, M. (2006). *WITCH. A World Induced Technical Change Hybrid Model* (Issue 2006_46). https://EconPapers.repec.org/RePEc:ven:wpaper:2006_46

Brandt, A. R. (2017). How Does Energy Resource Depletion Affect Prosperity? Mathematics of a Minimum Energy Return on Investment (EROI). *BioPhysical Economics and Resource Quality*, *2*(1). https://doi.org/10.1007/s41247-017-0019-y

Bundesanstalt für Geowissenschaften und Rohstoffe (BGR), 2017. Energy Study: Data and Developments Concerning German and Global Energy Supplies

Campbell, J. Y., & Mankiw, N. G. (1991). The response of consumption to income: A cross-country investigation. *European Economic Review*, *35*(4), 723–756. https://doi.org/https://doi.org/10.1016/0014-2921(91)90033-F

Capellán-Pérez, I. (2016). *Development and application of environmental integrated assessment modelling towards sustainability*. University of the Basque Country.

Capellán-Pérez, I., Arto, I., Polanco-Martínez, J. M., González-Eguino, M., & Neumann, M. B. (2016). Likelihood of climate change pathways under uncertainty on fossil fuel resource availability. *Energy & Environmental Science*, *9*(8), 2482–2496. https://doi.org/10.1039/C6EE01008C

Capellán-Pérez, I., de Blas, I., Nieto, J., de Castro, C., Miguel, L. J., Carpintero, Ó., Mediavilla, M., Lobejón, L. F., Ferreras-Alonso, N., Rodrigo, P., Frechoso, F., & Álvarez-Antelo, D. (2020). MEDEAS: a new modeling framework integrating global biophysical and socioeconomic constraints. *Energy & Environmental Science*, *13*(3), 986–1017. https://doi.org/10.1039/C9EE02627D

Capellán-Pérez, I., de Castro, C., & Miguel González, L. J. (2019). Dynamic Energy Return on Energy Investment (EROI) and material requirements in scenarios of global transition to renewable energies. *Energy Strategy Reviews*, *26*, 100399. https://doi.org/https://doi.org/10.1016/j.esr.2019.100399

Carroll, C. D. (1996). Buffer-Stock Saving and the Life Cycle/Permanent Income Hypothesis. *National Bureau of Economic Research Working Paper Series*, *No. 5788*. https://doi.org/10.3386/w5788

Carroll, C., & Kimball, M. (1995). *On the Concavity of the Consumption Function*. https://EconPapers.repec.org/RePEc:wpa:wuwpma:9503003

Commission, E., Eurostat, for Economic Co-operation, O., & Development. (2015). *ISCED 2011 operational manual : guidelines for classifying national education programmes and related qualifications*. Unesco. https://doi.org/doi/10.1787/978-92-64-228368-en

Creative Commons. (n.d.). *About The Licenses*. https://creativecommons.org/licenses/

de Blas, I., Miguel, L. J., & Capellán-Pérez, I. (2019). Modelling of sectoral energy demand through energy intensities in MEDEAS integrated assessment model. *Energy Strategy Reviews*, *26*(September 2018), 100419. https://doi.org/10.1016/j.esr.2019.100419

de Castro, C., & Capellán-Pérez, I. (2020). Standard, Point of Use, and Extended Energy Return on Energy Invested (EROI) from Comprehensive Material Requirements of Present Global Wind, Solar, and Hydro Power Technologies. *Energies 2020, Vol. 13, Page 3036*, *13*(12), 3036. https://doi.org/10.3390/EN13123036

De Castro, C., Mediavilla, M., Miguel, L. J., & Frechoso, F. (2011). Global wind power potential: Physical and technological limits. *Energy Policy*. https://doi.org/10.1016/j.enpol.2011.06.027

Després, J., Keramidas, K., Schmitz, A., & Kitous, A. (2018). POLES-JRC model documentation. *Joint Research Centre*. https://doi.org/10.2760/814959

Diaz, D., & Moore, F. (2017). Quantifying the economic risks of climate change. *Nature Climate Change*, *7*(11), 774–782. https://doi.org/10.1038/nclimate3411

Dupont, E., Koppelaar, R., & Jeanmart, H. (2018). Global available wind energy with physical and energy return on investment constraints. *Applied Energy*, *209*(July 2017), 322–338. https://doi.org/10.1016/j.apenergy.2017.09.085

Dupont, E., Koppelaar, R., & Jeanmart, H. (2020). Global available solar energy under physical and energy return on investment constraints. *Applied Energy*, *257*. https://doi.org/10.1016/j.apenergy.2019.113968

Edmonds, J., & Reilly, J. M. (1985). Global energy: Assessing the future. *Oxford University Press*. https://doi.org/https://doi.org/10.1016/0140-9883(86)90007-1

Eggertsson, G. B., & Krugman, P. (2012). Debt, Deleveraging, and the Liquidity Trap: A Fisher-Minsky-Koo Approach*. *The Quarterly Journal of Economics*, *127*(3), 1469–1513. https://doi.org/10.1093/qje/qjs023

Energy Watch Group. (2006). *URANIUM RESOURCES AND NUCLEAR ENERGY*.

Energy Watch Group. (2013). Fossil and Nuclear Fuels – the Supply Outlook-Boelkow-Foundation /Reiner- Lemoine-Foundation. *EWG-Series*, *March*. www.energywatchgroup.org

European Commission. (n.d.). *EDGAR - The Emissions Database for Global Atmospheric Research*. Retrieved July 28, 2023, from https://edgar.jrc.ec.europa.eu/emissions_data_and_maps

Eurostat. (2015). *ESS guidelines on temporal disaggregation : benchmarking and reconciliation*.

Eurostat. (2019). *Glossary:Metadata*. https://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:Metadata

Evans, S., & Housefather, Z. (2020, April 18). *Q&A: How 'integrated assessment models' are used to study climate change*.

Fiddaman, T., Siegel, L. S., Sawin, E., Jones, A. P., & Sterman, J. (2018). *C-ROADS SIMULATOR R EFERENCE GUIDE v83*. *March*.

Hall, C., Balogh, S., & Murphy, D. (2009). What is the Minimum EROI that a Sustainable Society Must Have? *Energies*, *2*(1), 25–47. https://doi.org/10.3390/en20100025

Hardt, L., & O'Neill, D. W. (2017). Ecological Macroeconomic Models: Assessing Current Developments. *Ecological Economics*, *134*, 198–211. https://doi.org/https://doi.org/10.1016/j.ecolecon.2016.12.027

Head, K., & Mayer, T. (2013). *Gravity Equations: Workhorse,Toolkit, and Cookbook*.

Hertwich, E. G., Gibon, T., Bouman, E. A., Arvesen, A., Suh, S., Heath, G. A., Bergesen, J. D., Ramirez, A., Vega, M. I., & Shi, L. (2015). Integrated life-cycle assessment of electricity-supply scenarios confirms global environmental benefit of low-carbon technologies. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(20), 6277–6282. https://doi.org/10.1073/pnas.1312753111

Hope, C. (2011). *The Social Cost of CO2 from the PAGE09 Model*.

Houghton, J., & Siegel, M. (2015). Advanced data analytics for system dynamics models using PySD. *Proceedings of the 33th International Conference of the System Dynamics Society*, *3*(4).

IEA. (2019). IEA World Energy Statistics and Balances. In *World Energy Statistics and Balances (database)*. IEA/OECD.

IRENA. (2022). *IRENA Data & Statistics*. https://pxweb.irena.org/pxweb/en/IRENASTAT/IRENASTAT__Power Capacity and Generation/ELECCAP_2022_cycle2.px/

Jappelli, T., & Pistaferri, L. (2014). Fiscal Policy and MPC Heterogeneity. *American Economic Journal: Macroeconomics*, *6*(4), 107–136. https://doi.org/10.1257/mac.6.4.107

Jappelli, T., & Pistaferri, L. (2020). Reported MPC and Unobserved Heterogeneity. *American Economic Journal: Economic Policy*, *12*(4), 275–297. https://doi.org/10.1257/pol.20180420

Jiang, Y., Yan, D., Guo, S., Hu, S., 2018. China Building Energy Use 2018.

José Luengo-Prado, M. (2006). Durables, nondurables, down payments and consumption excesses. *Journal of Monetary Economics*, *53*(7), 1509–1539. https://doi.org/https://doi.org/10.1016/j.jmoneco.2005.05.010

Khalili, S., Rantanen, E., Bogdanov, D., & Breyer, C. (2019). Global Transportation Demand Development with Impacts on the Energy Demand and Greenhouse Gas Emissions in a Climate-Constrained World. *Energies 2019, Vol. 12, Page 3870*, *12*(20), 3870. https://doi.org/10.3390/EN12203870

Kim, K., Kratena, K., & Hewings, G. J. D. (2013). *The Extended Econometric Input-output Model With Heterogeneous Household Demand System The Extended Econometric Input-output Model With Heterogeneous Household Demand System*. *217*. www.real.illinois.edu/

Kratena, K., Streicher, G., Salotti, S., Sommer, M., & Valderas Jaramillo, J. (2017). *FIDELIO 2: Overview and theoretical foundations of the second version of the Fully Interregional Dynamic Econometric Long-term Input- Output model for the EU-27*. https://doi.org/10.2760/313390

Kriegler, E., Riahi, K., Bauer, N., Schwanitz, V. J., Petermann, N., Bosetti, V., Marcucci, A., Otto, S., Paroussos, L., Rao, S., Arroyo Currás, T., Ashina, S., Bollen, J., Eom, J., Hamdi-Cherif, M., Longden, T., Kitous, A., Méjean, A., Sano, F., … Edenhofer, O. (2015). Making or breaking climate targets: The AMPERE study on staged accession scenarios for climate policy. *Technological Forecasting and Social Change*, *90*, 24–44. https://doi.org/https://doi.org/10.1016/j.techfore.2013.09.021

Laherrère, J., Hall, C.A.S., Bentley, R., 2022. How much oil remains for the world to produce? Comparing assessment methods, and separating fact from fiction. Curr. Res. Environ. Sustain. 4, 100174. https://doi.org/https://doi.org/10.1016/j.crsust.2022.100174.

Lentzner, H. (2010). *The effect of education on adult mortality and disability: a global perspective*. *2010*, 201–236. http://austriaca.at/?arp=0x0024fc45

Loulou, R., Remme, U., Kanudia, A., Lehtilä, A., & Goldstein, G. (2005). Documentation for the TIMES model. *Energy Technology Systems Ananlysis Programme (ETSAP)*.

Lund, H., Thellufsen, J. Z., Østergaard, P. A., Sorknæs, P., Skov, I. R., & Mathiesen, B. V. (2021). EnergyPLAN – Advanced analysis of smart energy systems. *Smart Energy*, *1*, 100007. https://doi.org/10.1016/j.segy.2021.100007

Lutz, W., & KC, S. (2010). Dimensions of global population projections: what do we know about future population trends and structures? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1554), 2779–2791. https://doi.org/10.1098/rstb.2010.0133

Lutz, W., & KC, S. (2011). Global Human Capital: Integrating Education and Population. *Science*, *333*(6042), 587–592. https://doi.org/10.1126/science.1206964

Luzzati T., Distefano T., Böck E., & Wergles N. (2021). *LOCOMOTION Deliverable 8.2: Policy measures and objectives selection and formulation*.

Manne, A. S., & Richels, R. G. (2005). Merge: An Integrated Assessment Model for Global Climate Change. In R. Loulou, J.-P. Waaub, & G. Zaccour (Eds.), *Energy and Environment* (pp. 175–189). Springer US. https://doi.org/10.1007/0-387-25352-1_7

Martin-Martinez, E., Samsó, R., Houghton, J., & Solé, J. (2022). PySD: System Dynamics Modeling in Python. *Journal of Open Source Software*, *7*(78), 4329. https://doi.org/10.21105/JOSS.04329

Meadows, D. H., Meadows, D. L., Randers, J., & Behrens, W. W. (1972). *The limits to growth*.

Messner, S., & Strubegger, M. (1995). *User's Guide for MESSAGE III*.

Nations, U. (1995). *UNITED NATIONS STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE CONFERENCE OF EUROPEAN STATISTICIANS METHODOLOGICAL MATERIAL GUIDELINES FOR THE MODELING OF STATISTICAL DATA AND METADATA*.

Nordhaus, W. D. (1979). *Efficient use of energy resources*. Yale University Press,New Haven, CT. https://www.osti.gov/biblio/5246359

Ólafsdóttir, A. H., Capellán-Pérez, I., Sverdrup, H. U., Ploiner, C., Goritschnig, W., Eggler, L., De Castro, C., Mediavilla, M., Álvarez Antelo, D., Pulido, D., Manzanera, G., de Blas, I., Frechoso, F., & Arto, I. (n.d.). LOCOMOTION Deliverable 7.3: Models of Energy technologies. *2021*. Retrieved July 24, 2023, from https://www.locomotion-h2020.eu/resources/main-project-reports/

Olafsdottir, A. H., & Sverdrup, H. U. (2021). Modelling Global Nickel Mining, Supply, Recycling, Stocks-in-Use and Price Under Different Resources and Demand Assumptions for 1850–2200. Mining, Metallurgy and Exploration, 38(2), 819–840. https://doi.org/10.1007/S42461-020-00370-Y

O'Neill, B. C., Kriegler, E., Riahi, K., Ebi, K. L., Hallegatte, S., Carter, T. R., Mathur, R., & van Vuuren, D. P. (2014). A new scenario framework for climate change research: the concept of shared socioeconomic pathways. *Climatic Change*, *122*(3), 387–400. https://doi.org/10.1007/s10584-013-0905-2

Open source initiative. (n.d.). *The MIT License*. https://opensource.org/license/mit/

Papagianni, S., Adam, A., Olafsdottir, A. H., Sverdrup, H. U., Samsó, R., Bjelic, I. B., Capellán-Pérez, I., & Castro, C. de. (2020). *LOCOMOTION Deliverable 7.2 Module of energy resources availability*. *November*. https://www.locomotion-h2020.eu/resources/main-project-reports/

Parrado-Hernando, G., Herc, L., Feijoo-Palacios, F. A., & Capellán-Pérez, I. (2023). Capturing Features of Hourly-resolution Energy Models in an Integrated Assessment Model. In SDEWES (Ed.), *18th SDEWES Conference Dubrovnik 2023*.

Pulido Sánchez, D., Capellán Pérez, I., Mediavilla Pascual, M., de Castro Carranza, C., & Frechoso Escudero, F. A. (2021). ANALYSIS OF THE MATERIAL REQUIREMENTS OF GLOBAL EL... - DYNA JOURNAL ENGINEERING. *DYNA Ingeniería e Industria*, *96*, 207–213. https://doi.org/https://doi-org.ponton.uva.es/10.6036/9893

Pulido Sánchez, D., Capellán-Pérez, I., & Mediavilla Pascual, M. (2020). *Análisis de los requerimientos materiales de la transición hacia una movilidad eléctrica*. http://uvadoc.uva.es/handle/10324/41646

Pulido-Sánchez, D. (2022). *Requerimientos materiales y EROI de las tecnologías fotovoltaicas en la transición energética*. https://uvadoc.uva.es/handle/10324/56539

Pulido-Sánchez, D., Capellán-Pérez, I., de Castro, C., & Frechoso, F. (2022). Material and energy requirements of transport electrification. *Energy & Environmental Science*, *15*(12), 4872–4910. https://doi.org/10.1039/D2EE00802E

Rogner, H.-H., Aguilera, R. F., Archer, C., Bertani, R., Bhattacharya, S. C., Dusseault, M. B., Gagnon, L., Haberl, H., Hoogwijk, M., Johnson, A., Rogner, M. L., Wagner, H., & Yakushev, V. (2012). Energy and Economy. In Global Energy Assessment - Toward a Sustainable Future. In *Global Energy Assessment - toward a Sustainable Future* (Cambridge, pp. 385–421).

Rogner, H.-H., Aguilera, R. F., Archer, C. L. C., Bertani, R., Bhattacharya, S. C., Dusseault, M. B., Gagnon, L., Haberl, H., Hoogwijk, M., Johnson, A., Rogner, M. L., Wagner, H., Yakushev, V., Arent, D. J., Bryden, I., Krausmann, F., Odell, P., Schillings, C., Shafiei, A., & Zou, J. (2012). Energy Resources and Potentials. In J. Zou (Ed.), *Global Energy Assessment - Toward a Sustainable Future* (pp. 423–512). Cambridge University Press. https://doi.org/10.1017/cbo9780511793677.013

Roson, R., & Sartori, M. (2016). Estimation of Climate Change Damage Functions for 140 Regions in the GTAP 9 Database. *Journal of Global Economic Analysis*, *1*(2). https://doi.org/10.21642/jgea.010202af

Rotmans, J., De Boois, H., & Swart, R. J. (1990). An integrated model for the assessment of the greenhouse effect: The Dutch approach. *Climatic Change*, *16*(3), 331–356. https://doi.org/10.1007/BF00144508

Sarantis, N., & Stewart, C. (2003). Liquidity constraints, precautionary saving and aggregate consumption: an international comparison. *Economic Modelling*, *20*(6), 1151–1173. https://doi.org/https://doi.org/10.1016/S0264-9993(02)00080-9

Schaphoff, S., Von Bloh, W., Rammig, A., Thonicke, K., Biemans, H., Forkel, M., Gerten, D., Heinke, J., Jägermeyr, J., Knauer, J., Langerwisch, F., Lucht, W., Müller, C., Rolinski, S., & Waha, K. (2018). LPJmL4 - A dynamic global vegetation model with managed land - Part 1: Model description. *Geoscientific Model Development*, *11*(4), 1343–1375. https://doi.org/10.5194/gmd-11-1343-2018

Scrieciu, S., Rezai, A., & Mechler, R. (2013). On the economic foundations of green growth discourses: the case of climate change mitigation and macroeconomic dynamics in economic modeling. *WIREs Energy and Environment*, *2*(3), 251–268. https://doi.org/https://doi.org/10.1002/wene.57

Sommer, M., & Kratena, K. (2017). The Carbon Footprint of European Households and Income Distribution. *Ecological Economics*, *136*, 62–72. https://doi.org/https://doi.org/10.1016/j.ecolecon.2016.12.008

Stadler, K., Wood, R., Bulavskaya, T., Södersten, C.-J., Simas, M., Schmidt, S., & Tukker, A. (2019). *EXIOBASE 3 (Version 3.8.1) [Data set]*. Zenodo.

Stehrer, R., Bykova, A., Jäger, K., Reiter, O., & Schwarzhappel, M. (2019). *Industry Level Growth and Productivity Data with Special Focus on Intangible Assets The Vienna Institute for International Economic Studies Wiener Institut für Internationale Wirtschaftsvergleiche*. www.wiiw.ac.at

Sterman, J. D. (2000). *Business dynamics: systems thinking and modeling for a complex world*.

Sverdrup, Harald U., Ragnarsdottir, K. V., Koca, D., Vala, K., & Koca, D. (2014). On modelling the global copper mining rates, market supply, copper price and the end of copper reserves. *Resources, Conservation and Recycling*, *87*, 158–174. https://doi.org/10.1016/j.resconrec.2014.03.007

Sverdrup, H., & Ragnarsdóttir, K. V. (2014a). Natural Resources in a Planetary Perspective. *Geochemical Perspectives*, *3*(2), 129–341. https://doi.org/10.7185/geochempersp.3.2

Sverdrup, Harald U., Ragnarsdottir, K. V., & Koca, D. (2015). Aluminium for the future: Modelling the global production, market supply, demand, price and long term development of the global reserves. *Resources, Conservation and Recycling*, *103*, 139–154. https://doi.org/10.1016/j.resconrec.2015.06.008

Sverdrup, Harald U., Ragnarsdottir, K. V., & Koca, D. (2017). An assessment of metal supply sustainability as an input to policy: security of supply extraction rates, stocks-in-use, recycling, and risk of scarcity. *Journal of Cleaner Production*, *140*, 359–372. https://doi.org/10.1016/j.jclepro.2015.06.085

Sverdrup, Harald Ulrik, & Olafsdottir, A. H. (2019). Assessing the Long-Term Global Sustainability of the Production and Supply for Stainless Steel. *BioPhysical Economics and Resource Quality*, *4*(2). https://doi.org/10.1007/s41247-019-0056-9

Sverdrup, Harald Ulrik, Olafsdottir, A. H., & Ragnarsdottir, K. V. (2019). On the long-term sustainability of copper, zinc and lead supply, using a system dynamics model. *Resources, Conservation and Recycling: X*, *4*(December 2018), 100007. https://doi.org/10.1016/j.rcrx.2019.100007

United Nations. (2010). *System of National Accounts 2008*. United Nations. https://doi.org/10.18356/4fa11624-en

United Nations. (2023). *United Nations database*. https://data.un.org

United Nations Development Programme. (2020). *Human development report 2020: the next frontier: human development and the anthropocene.* (2020th ed.).

US EIA, & EIA. (2020). *Annual Energy Outlook 2020 with Projections to 2050. (No. AEO2020). U.S. Energy Information Administration*. https://www.eia.gov/outlooks/aeo/pdf/AEO2020.pdf

V Pastor, A., Capela Lourenço, T., Samsóm, R., Martín, E., Ferreras-Alonso, N., Ramos, I., Papagianni, S., Mediavilla, M., Miguel González, L. J., Parrado-Hernando, G., Reyes, L., & Oakes, R. (2021). *LOCOMOTION Deliverable 6.3: Climate change impacts and adaptation module*. Delliverable 6.3 LOCOMOTION. https://www.researchgate.net/publication/353286611_Climate_change_impacts_and_adaptation_module

van de Ven, D.-J., González-Eguino, M., & Arto, I. (2018). The potential of behavioural change for climate change mitigation: a case study for the European Union. *Mitigation and Adaptation Strategies for Global Change*, *23*(6), 853–886. https://doi.org/10.1007/s11027-017-9763-y

van den Berg, N. J., Hof, A. F., Akenji, L., Edelenbosch, O. Y., van Sluisveld, M. A. E., Timmer, V. J., & van Vuuren, D. P. (2019). Improved modelling of lifestyle changes in Integrated Assessment Models: Cross-disciplinary insights from methodologies and theories. *Energy Strategy Reviews*, *26*, 100420. https://doi.org/https://doi.org/10.1016/j.esr.2019.100420

Weyant, J. (2017). Some Contributions of Integrated Assessment Models of Global Climate Change. *Review of Environmental Economics and Policy*, *11*(1), 115–137. https://doi.org/10.1093/reep/rew018

Wang, J., & Bentley, Y. (2020). Modelling world natural gas production. Energy Reports, 6, 1363–1372. https://doi.org/10.1016/j.egyr.2020.05.018

Wilkinson, M. D. (2016). *wilkinson-2016-the_FAIR_guiding_principles*. 1–9.

Zittel, W. (2012). *Feasible Futures for the Common Good. Energy Transition Paths in a Period of Increasing Resource Scarcities Progress Report 1: Assessment of Fossil Fuels Availability (Task 2a) and of Key Metals Availability (Task 2 b)*.

ANNEX

1.1. LIST OF CHANGES BETWEEN PYSD VERSIONS 0.11.0 AND 2.2.4

PySD 0.11.0 (18/12/2020)

We have taken as an initial version the following pull request https://github.com/julienmalard/pysd/pull/1. Main changes (new external.py, build.py, vensim2py.py, utils.py, functions.py, __init__.py, new requirement?):

- Remove duplicated functions as _col_to_num/str_to_int and _num_to_col/to_ABC.
- Documentation of functions that were not documented.
- Move all the external file reading functions to external.py file to clean functions.py file.
- Correction of cachetype: run for constant, step for data and no-cache for lookups.
- Creation of a new class External (initialized before Stateful objects), which avoids having to compute the time step of external data objects (which would slow down the models). Cleaning of the open files in Excel class after initialization.
- Implementation of multidimensional LOOKUPS and DATA (with the new add method we only need to do the interpolation one time for each matrix).

- Avoid computing a lot of objects to the same CONSTANT; DATA or LOOKUPS matrix, a unique object is created, and new information will be added with the new add method. Then the initialization of the external object will give the final object with all the dimensions.
- Corrections and simplification in the data reading functions.
- Some improvements when reading missing values.
- Giving more description for some Errors.
- Reading data by cell names using OpenPyXL, only available for python >= 3.6.
- Dimensions are now passed to the objects to preserve the order. This is necessary for compatibility with python < 3.7, where the dictionaries are not ordered.
- Corrected subscripted Delays to make them work in all the cases.
- Corrected operations between subscripted variables such as SUM(a[object!]*b[object!]).
- Correct subsetting data in the outputs by coordinate values and not by a coords dictionary (makes Vensim® comparison work when there is more than one dimension with the same coordinate values).
- Added several tests for all the new functions.
- Added several new model tests.
- Simplify some tests that were taking too long.
- Wrapping the subscripted variables that have different arguments than the original ones with a function that takes the subscripts as arguments (utils.rearrange) . This makes possible dimension name switching and transpositions. It also makes future implementation of subsetting easier.
- Making a subscript decorator for up-dimensioning and reordering data (need to add new cache to make it work).
- Added support for user interaction with subscripted variables (similar to the interaction with scalars). Add an explanation of how to do that in the documentation.
- Made most of the new implementations work on Python 2.7 and 3.5.
- Subscripted if_then_else, xidz, zidz support.
- Cache functions have been moved to cache.py file and converted to a class which makes it easier to manage the cache. Now the cache decorators are @cache.run and @cache.step to avoid double wrapping and make the code simpler. However, @cache("horizon") will still work.
- Now we can use the clean method with no need to iterate over all the elements of the model as a nested dictionary is used to save cache values. The cache step can be cleaned using reset method which is called in the integration (we avoid checking the cache value time when calling a function).
- The __call__ method, ensures backward compatibility (removed in version 1.0.0)

PySD 1.0.0 (08/01/2021)

General cleaning of PySD library, dropping support for Python 2 and Python < 3.7. Several backward incompatible changes made:

- General cleanup, removing dropped methods. This will make developing new features easier from now on.
- Lazy evaluation of if then else implemented. This will save a lot of time during the computation as the regular evaluation of if then else in pymedeas was taking a lot of time.
- Intelligent importing of PySD functions and methods in the model file. This makes the model file look much simpler and easier to follow.

PySD 1.1.0 (11/02/2021)

- Support for subscript subrange, already existing test model working.
- Issue 216 solved, test model added.
- Solved some bugs when reading missing values from excel files, test model added. User can now decide if they want to raise an error, warning or ignore any message.
- Big changes in excel files reading methodology to make it work more than 90% faster with big excel files models.
- Added several tests

PySD 1.2.0 (05/05/2021)

- Added a progress bar during integration which can be activated passing the optional argument progress=True to the run function. progressbar2 library used https://pypi.org/project/progressbar2/
- Corrected bug when subsetting dimensions due to the shared coordinate name with other dimensions. Corrected by avoiding using dimension name when subsetting.
- Correct bugs of merging coordinates from subranges.
- Clean vensim2py.
- Correct some bugs of visitation calls and lookups with subscripts.
- Add option to include functions still no added on PySD on the parser and continue translating model file.
- Clean models' files: rearrange is called only when needed and constant xarrays are initialized with a float.
- Add an initialization argument when loading the model.
- Add support for calling external lookups with subscripted arguments with several dimensions.
- More descriptive errors when Vensim® parser fails.
- Added several tests.
- Tests models added.

PySD 1.3.0 (11/05/2021)

- Add support for Vensim® Sample If True function (https://www.vensim.com/documentation/fn_sample_if_true.html)
- Add support for reading subscript by numeric range (https://www.vensim.com/documentation/22090.html)

PySD 1.3.1 (17/05/2021)

- Minor bug fix.
- Added the possibility of keeping missing values in External objects.

PySD 1.4.0 (08/06/2021)

- Add support for subscripted smooth and trend.
- Improvements in delays solving old tests (test_delays/test_delay_pipelines) and new test (test_delay_fixed).
- Solve user interaction errors when trying to assign values to a lookup variable.
- Add identifiers to stateful objects to make debugging easier.
- Add raises with stateful identifiers.
- Test added.
- Test models added.

- Solves issue https://github.com/JamesPHoughton/pysd/issues/226.

PySD 1.5.0 (30/06/2021)

- Move test_utils.py functions to pysd/tools/benchmarking.py and document them to let users use them.
- Add support for logical operations between subscripted variables.
- Add support for logical operations without parenthesis.
- Write outputs directly in a Dataframe using the timestamp.
- Write the values of the constant variables after the integration.
- Add an option to totally flatten an output Dataframe.
- Solve some bugs related to subscripts.
- Add cache to params arguments.
- Write some tests in more-tests instead of having them hardcoded in the test files.
- Solve bugs related to External._resolve_file.
- Move builder import manager to a class to make it more strong and easier to manage.
- Include _root in the python model file only if the model has external objects.
- Add the possibility to select the as default columns only the ones with cache 'step'.
- Totally subsetted subscripted variables converted to floats.
- Add tests and test model.
- Solves:
    - https://github.com/JamesPHoughton/pysd/issues/248
    - https://github.com/JamesPHoughton/pysd/issues/69
    - https://github.com/JamesPHoughton/pysd/issues/268
    - https://github.com/JamesPHoughton/pysd/issues/56

PySD 1.6.0 (02/07/2021)

- Add command line interface to call PySD from command line.
- Improve setting initial conditions for all stateful elements.
- Add possibility to export model current state and import it later in other session.
- Solve bugs related to SampleIfTrue.
- Correct bugs related to initial time.
- Update documentation.
- Add tests.
- Solves:
    - https://github.com/JamesPHoughton/pysd/issues/271
    - https://github.com/JamesPHoughton/pysd/issues/273

PySD 1.6.1 (02/07/2021)

- Add requirements.txt and README.md to MANIFEST.in in order to make it possible to build the package in conda-forge.

PySD 1.7.0 (02/07/2021)

- Add possibility to export to a pickle the states and other needed values of the stateful objects using.
- Add possibility to use exported values of stateful objects as initial conditions.

- Add control variables to the run arguments, final_time, time_step and saveper can be modified easily when calling run.
- A bug in the SampleIfTrue function was solved.
- Solve initialization order bug.

PySD 1.8.0 (07/07/2021)

- Putting model variables in separate modules according to the views they belong to in the Vensim® model.
- Add tests to check the new functionality.
- Update the documentation to describe the new functionality.
- LICENSE added to MANIFEST as suggested by conda-forge maintainer.
- Added conda version badge to README and installation instructions to docs.
- Splitting option working with CLI as well.
- Add some tests for subscripts range parsing error.

PySD 1.8.1 (13/07/2021)

- Add suport for dynamical final time, stopping the model automatically.
- Correct bugs when subsetting ranges.
- Close open xlsx files.
- Test model added.

PySD 1.9.0 (09/08/2021)

- Divide the python model into folders and files depending on the Vensim® view and subview.
- Add support for INVERT MATRIX.
- Added support for logical comparisons between subscript ranges.
- Improvements in benchmarking tools.
- Add tests and test models.
- Solves:
    - https://github.com/JamesPHoughton/pysd/issues/285
    - https://github.com/JamesPHoughton/pysd/issues/286

PySD 1.10.0 (10/09/2021)

- Fix bugs related to merging subscripts and subscripts subranges.

PySD 1.11.0 (07/10/2021)

- Add support for Forecast.
- Add support for: NA.
- Add support for ELMCOUNT.
- Allow any level of submodules.
- Allow passing subscript ranges as function arguments.
- Correct bugs when reading external data with no values.
- Add test model.
- Add tests.
- Update docs.
- Solves:

- https://github.com/JamesPHoughton/pysd/issues/264
- https://github.com/JamesPHoughton/pysd/issues/154 (for Vensim® only)
- https://github.com/JamesPHoughton/pysd/issues/294
- https://github.com/JamesPHoughton/pysd/issues/295

PySD 2.0.0 (08/11/2021)

- Improving caching strategy.
- Add dependencies dictionary to translated models.
- Improves in time management allowing dynamic control variables.
- Improves in model components management.
- Add __getitem__ method to pysd model to easily get the values of a component.
- Add get_series_data method to pysd model to get the values for all interpolation stamps of Data and Lookups.
- Add tests.
- Solves:
    - https://github.com/JamesPHoughton/pysd/issues/115
    - https://github.com/JamesPHoughton/pysd/issues/238
    - https://github.com/JamesPHoughton/pysd/issues/298

PySD 2.1.0 (13/12/2021)

- Allow working with Vensim® models with regular DATA declaration, reading the data from a csv or tab file when loading the model. Giving several options to the user.
- Allow comparing two files with missing values and return True if the missing values are at the same place.
- Allow running submodel or the current model, selecting it by variable names or modules (produced when translating a file per views).
- Bug fixes.
- Improve and update documentation.
- Add models.

PySD 2.2.0 (16/12/2021)

- Enable using encoding information for reading Vensim® model files.
- Read modules files with "UTF-8" encoding.
- Correct the paths and use pathlib.Path type object in the model file for external objects, modules and Macros.
- Correct several tests to make them run on Windows also.
- Add windows to GitHub Actions CI.
- Solve dependencies with active initial objects.
- Solves:
    - https://github.com/JamesPHoughton/pysd/issues/305
    - https://github.com/JamesPHoughton/pysd/issues/303

PySD 2.2.1 (07/02/2022)

- Solve bugs of arrays with line breaks.

PySD 2.2.2 (11/03/2022)

- Solve missing smile grammar in packages.

PySD 2.2.3 (17/02/2022)

- Solve bug when converting to floats 0-d dataarrays.

PySD 2.2.4 (24/04/2022)

- Update progressbar dependency to progressbar2.

Since PySD 3.0.0, the new changes are reported in the *What's new* section of the documentation (https://pysd.readthedocs.io/en/master/whats_new.html).