iDesignRES

Building Stock Energy Model: Complementary documentation to the model code

WP1 – Multi-Physics component models: Implementation and development

Task 1.4: Energy consumers: industry, **buildings**, transport

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

Work Package	WP1 Multi-Physics component models: Implementation and development
Dissemination Level	Public
Author(s)	Eneko Arrizabalaga, Nekane Hermoso, Ricardo Alonso, Ainhoa Pedera, Alberto Molinuevo
Primary Contact and Email	eneko.arrizabalaga@tecnalia.com

This document has been prepared in the framework of the European project iDesignRES. This project has received funding from the European Union's Horizon Europe programme under grant agreement no. 101095849.

1. Table of Content

1. BUILDING STOCK ENERGY MODEL - PURPOSE OF THE MODEL....4

2. MODEL DESIGN PHILOSOPHY....4

3. INPUT TO AND OUTPUT FROM THE MODEL....6

4. IMPLEMENTED FEATURES....13

5. CORE ASSUMPTIONS....15

5.1. General assumptions....15
5.2. Mathematical description....15

- Pre-processing of basic data of the building stock....16
- Pre-processing of input data used for the solar process....21
- Calculation of energy needs and energy consumption of buildings in the region....23

1. Building Stock Energy Model - Purpose of the model

The objective of the model is to simulate the energy performance of the building stock of any region in Europe (NUTS Level2) both for an initial diagnosis and to evaluate different years of the transition period considered in the scenarios proposed. The aim is to cover the building stock for different uses including the residential and tertiary sector with a high degree of disaggregation in both cases.

Besides, the model aims to generate information that provides greater granularity to the building sector models generated in the project for higher scales such as the national scale.

2. Model design philosophy

In the case of the building sector, when developing new models and functionalities that allow ESM type models (traditionally focused on covering the European and country scales) to reach a regional resolution, it is necessary to find the balance between the agility of calculation and this greater detail of analysis. In this process, it is worth highlighting the potential associated with the use of georeferenced information to capture the specificities of each region in terms of building typologies. This allows a more detailed disaggregation of the building energy model at the country level, capturing aspects such as the number of buildings, surface area, age, or the most specific use for each building typology in the region, as well as their geographic distribution. Traditional methods used for obtaining and processing detailed georeferenced data are applied at smaller scales such as the district or urban scales since they are based on cadastral data. Moreover, they are not easily replicable and automatable for other cities and regions due to their high heterogeneity. This aspect related to the specificity to the case study and the complexity of data preprocessing is an important barrier that makes it unusual to disaggregate national models at the regional level with a high degree of disaggregation of building typologies based on bottom-up data collected.

The developed model aims to help break this barrier by treating and preprocessing geometric and semantic information of each of the buildings (level of detail at the building portal level) in the region. It follows this bottom-up georeferenced information processing approach but starting from georeferenced data available for the whole Europe and developed ad-hoc to cover larger scales such as the regional scale. This information is then used to adapt the energy calculation of the building sector through the use of building archetypes that provide greater detail for each building use. The model performs hourly simulations for given years so that they can be used for initial diagnosis and analysis of potential future scenarios.

The Figure 1 shows the complete modeling framework of the regional building stock energy model of iDesignRES project.

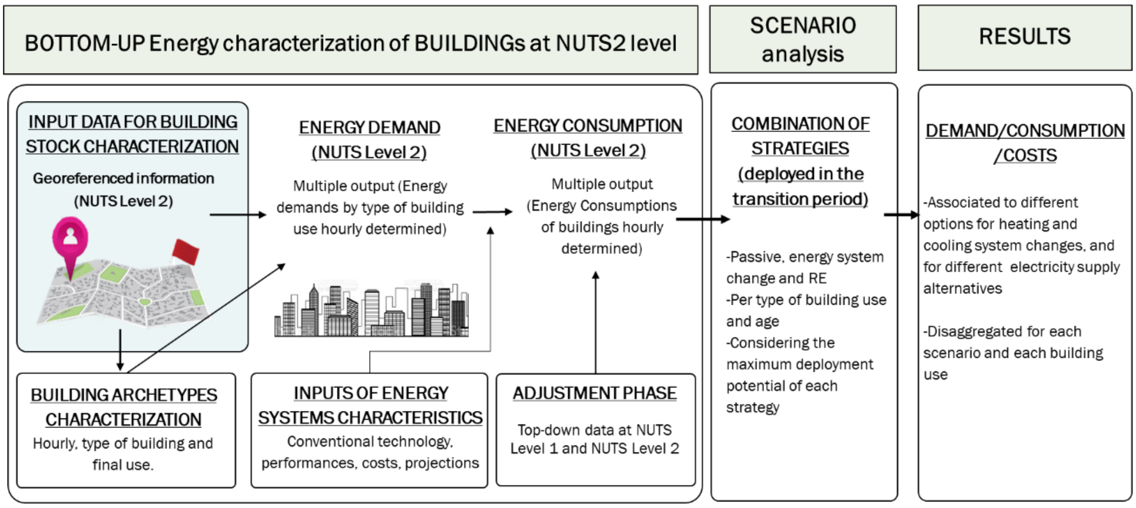

Figure 1. Modelling framework of the regional building stock energy model of iDesignRES project.

The model can be divided into three main blocks:

- Characterization and initial diagnosis of the building stock in the selected region (including simulation of energy demand and consumption).
- Configuration and simulation of scenarios
- Obtaining results

Going into more detail in the case of the first block (diagnosis), this has been divided in turn into two large blocks of analysis due to the large volume of processes and data processing required in the field of building stock energy models.

- Preprocessing of basic data of the building stock:
    - The model generates in a first instance the most relevant information required as input to the energy model in an automated way for the selected NUTS Level 2 region of Europe. Avoiding the user having to contact the different local, provincial, and regional entities in question for the request of cadaster shape files. The characterization obtained is at the portal level of each building in the region collecting the main characteristics in terms of use, sub-use, age, geometry, area and height, as well as their geographical distribution that can be used for finer analysis than NUTS Level 2.
    - In a second simulation phase, the model treats the information obtained and simplifies it to the level of archetypes of representative building typologies (according to the form factor) to be simulated for the region.
- Calculation of energy needs and energy consumption of buildings in the region.

3. Input to and output from the model

<u>Input data:</u>

(a) For initial diagnosis: energy demand (base year assessment).

- Data to be filled in by the user: NUTS Level 2 Code or NUTS Level 3 Code.

REGION_SELECTION['NUTS_Level'] = nuts_level

REGION_SELECTION['Selected_Region'] = region_code

- Main data used by the model as input (**provided by the model** with the option to be modified by the user in case more accurate data are available): Meteorological data (Outdoor dry bulb temperature (hourly), Radiation, Solar gains, heating, and cooling periods); geometry, surface area, age, use of each building. Other building parameters (U-values, H/C base temperature, Window-to-wall ratio, adjacent buildings, etc.)

(b) For initial diagnosis: energy consumption calculation (base year assessment).

<u>For the building energy consumption analysis</u>

One of the main inputs for this part of the code is precisely the output of the preprocessing module described in section (a). Thus, for each building in the selected region the following parameters would be used as inputs.

```
BUILDING_STOCK_PREPROCESSED = [
  'Building_ID',
  'Use',
  'Age',
  'Footprint_Area',
  'Number_of_floors',
  'Volume',
  'Gross_floor_area',
  'Total_External_Facade_area',
  'Height'
]
```

On the other hand, main data used by the model as input (provided by the model with the option to be modified by the user in case more accurate data are available): Shares of fuels/technologies by building type (From statistical data or model results of a higher scale, used for the adjustment), Hourly profiles (for Heating, Cooling, DHW, Occupancy, Lighting, Equipment, kitchen), Installed power (Lighting, Equipment), Equipment performance, Fuel cost, Environmental impact factors.

As an example, the structure for the input of "voluntary" data is shown in case the modeler has more precise data for the case of share of technologies for a building typology and a fixed year can be observed below.

```
SHARE_TECHNOLOGIES_RESIDENTIAL (year) = {
   'Space heating': {
      'Overall': 'Value', 'Solids': 'Value', 'LPG': 'Value', 'Diesel': 'Value',
      'Gas heat pumps': 'Value', 'Natural gas': 'Value', 'Biomass': 'Value',
      'Geothermal': 'Value', 'Distributed heat': 'Value',
      'Advanced electric heating': 'Value', 'Conventional electric heating': 'Value',
      'BioOil': 'Value', 'BioGas': 'Value', 'Hydrogen': 'Value'
   },
   'Electricity in circulation': 'Value',
   'Space cooling': {
      'Overall': 'Value', 'Gas heat pumps': 'Value',
      'Electric space cooling': 'Value'
   },
   'Water heating': {
      'Overall': 'Value', 'Solids': 'Value', 'LPG': 'Value',
      'Diesel': 'Value', 'Natural gas': 'Value',
      'Biomass': 'Value', 'Geothermal': 'Value',
      'Distributed heat': 'Value', 'Electricity': 'Value',
      'BioOil': 'Value', 'BioGas': 'Value',
      'Hydrogen':  'Value', 'Solar':  'Value'
   },
   'Cooking': {
      'Overall':  'Value', 'Solids':  'Value',
      'LPG':  'Value', 'Natural gas':  'Value',
      'Biomass':  'Value', 'Electricity':  'Value'
   },
   **{'Appliances & Lighting': {'Electricity':  'Value'}}
}
```

Similarly, for entering time profiles, the user can enter specific values for each building typology differentiating between weekdays, Saturdays, and Sundays for each energy end-use. Below is an example for a building typology and for a heating profile. The process converts this profile into an hourly profile for the whole year.

```
# Heating profile for apartment blocks on weekdays

heating_profile_weekday = {
   'Time': [
      '0:00', '1:00', '2:00', '3:00', '4:00',
      '5:00', '6:00', '7:00', '8:00', '9:00',
      '10:00', '11:00', '12:00', '13:00', '14:00',
      '15:00', '16:00', '17:00', '18:00', '19:00',
      '20:00', '21:00', '22:00', '23:00'
   ],
   'Heating': [
      'Value' for _ in range(24)  # Placeholder for heating values
   ]
}
```

For the solar module

Data to be filled in by the user: NUTS Level 2 Code or NUTS Level 3 Code.

REGION_SELECTION['NUTS_Level'] = nuts_level

REGION_SELECTION['Selected_Region'] = region_code

For the building integrated solar module, the data preprocessing that is initially executed provides the following results that are used as the main inputs in the energy calculation phase.


```
{
  "Region": "A1",
  "Centroid_X": 125.45,
  "Centroid_Y": 98.30,
  "Total_Area": 456.78,
  "Max_Radiation": 150.0,
  "Average_Radiation": 120.5,
  "Threshold": 75.0,
  "Area_m2": 500.0,
  "Median_Radiation": 130.0,
  "Median_Radiation_X": 126.1,
  "Median_Radiation_Y": 97.8
},
```


Besides, other input parameters of the solar photovoltaic module in buildings can be defined specifically for each building use considered. Below you can see the inputs parameters in more detail and the building categories disaggregation covered by the model for this module.


```
{
  "Apartment": {
    "area_total": 10000000,
    "power": null,
    "capex": null,
    "tilt": 30,
    "azimuth": 180,
    "loss": 14,
    "efficiency": 18,
    "tracking": 0,
    "use_factor": 100,
    "distribution": null
  },
  "Single": {
    "area_total": 10000000,
    "power": null,
    "capex": null,
    "tilt": 30,
    "azimuth": 180,
    "loss": 14,
    "efficiency": 18,
```

```
    "tracking": 0,
    "use_factor": 100,
    "distribution": null
  },
  "Education": {
    "area_total": 10000000,
    "power": null,
    "capex": null,
    "tilt": 30,
    "azimuth": 180,
    "loss": 14,
    "efficiency": 18,
    "tracking": 0,
    "use_factor": 100,
    "distribution": null
  },
```


The same input parameters apply in the same way to the rest of the building uses that correspond to:

"Health"
"Hotels"
"Offices"
"Sport"
"Trade"
"Other"
"Industrial"

(c) For the scenarios

Main data used by the model as input: Shares of fuels/technologies of the scenario to be simulated (if available from model results of a higher scale, used for the adjustment). Investments/measurement of amount of technology deployed in each sector/type of buildings. For example (for refurbishment, the percentage of the total m^2 refurbished, for solar PV total MW installed or associated investment, etc.).

More details about the scenario configuration can be found in the section *5.2.2 Calculation of energy needs and energy consumption of buildings in the region*.

<u>Output data:</u>

The modeling output allows to evaluate the energy performance of the building sector under different assumptions, both for the base year and for future situations depending on the scenario defined.

It should also be noted that the results can be obtained at the regional level (NUTS level 2) offering the possibility to perform simulations for each of the NUTS Level 3 that compose it, allowing to evaluate in greater detail the behavior of each area within the region taking into account its main particularities.

As for the outputs of each of the defined scales, these can be obtained in a more detailed way according to the disaggregation shown in previous sections. That is, by sector, sub-sector, building age, etc. The results are provided both aggregated on an annual basis and disaggregated on an hourly basis for the simulated year for the main parameters: energy needs, energy consumption, costs and CO2 emissions.

As an example, the model outputs for the energy needs for each building typology will have this form:

```
building_stock_energy_consumption = {
   "building_id": [
      "Apartment Block_1_Before1945",
      "Apartment Block_3_Before1945",
      "Apartment Block_2_Before1945",
      # ... (other building IDs)
      "Apartment Block_1_2000to2010"
   ],
   "heating_demand": {
      "values": [409180905.90, 179001462.80, 556146269.94, ...],
      "unit": "KWh"
   },
   "cooling_demand": {
      "values": [12039583.42, 4552452.21, 15019742.38, ...],
      "unit": "KWh"
   },
   "dhw_demand": {
      "values": [114854454.59, 64108075.92, 165095234.63, ...],
      "unit": "KWh"
   },
   "cooking_demand": {
      "values": [50848244.29, 28381860.48, 73090790.00, ...],
      "unit": "KWh"
   },
   "lighting_demand": {
      "values": [34340099.87, 19167543.28, 49361488.55, ...],
      "unit": "KWh"
   },
   "equipment_demand": {
      "values": [195741060.51, 109256387.26, 281364065.74, ...],
      "unit": "KWh"
   }
}
```

As an example of the hourly results, the model outputs for the energy cosnumption for each building typology will have this form:

```
building_type = "Apartment block"
date = "2024-01-01"
energy_consumption = {
   "timestamp": [
      "2024-01-01 00:00", "2024-01-01 01:00", ..., "2024-01-01 23:00"
   ],
   "heating": [
      0.0, 157749.9248, ..., 2245684.85
   ],
```

"cooling": [
0.0, 0.0, ..., 0.0
],
"dhw": [
64653.90799, 25861.5632, ..., 258615.632
],
"lighting": [
17432.99081, 14542.09284, ..., 39465.13749
],
"equipment": [
99369.31232, 82890.98415, ..., 224954.1467
],
"cooking": [
0.0, 0.0, ..., 28968.4656
]
}

And as mentioned, these results can be obtained for the different categories:

```
building_stock_energy_consumption = {
  "building_categories": [
    "Apartment block",
    "Single family - Terraced houses",
    "Hotels and Restaurants",
    "Health",
    "Education",
    "Offices",
    "Trade",
    "Sport",
    "Other non-residential buildings"
  ],
  "energy_end_uses": {
    "heating": {"unit": "KWh"},
    "cooling": {"unit": "KWh"},
    "dhw": {"unit": "KWh"},
    "lighting": {"unit": "KWh"},
    "equipment": {"unit": "KWh"},
    "cooking": {"unit": "KWh"}
  }
}
```

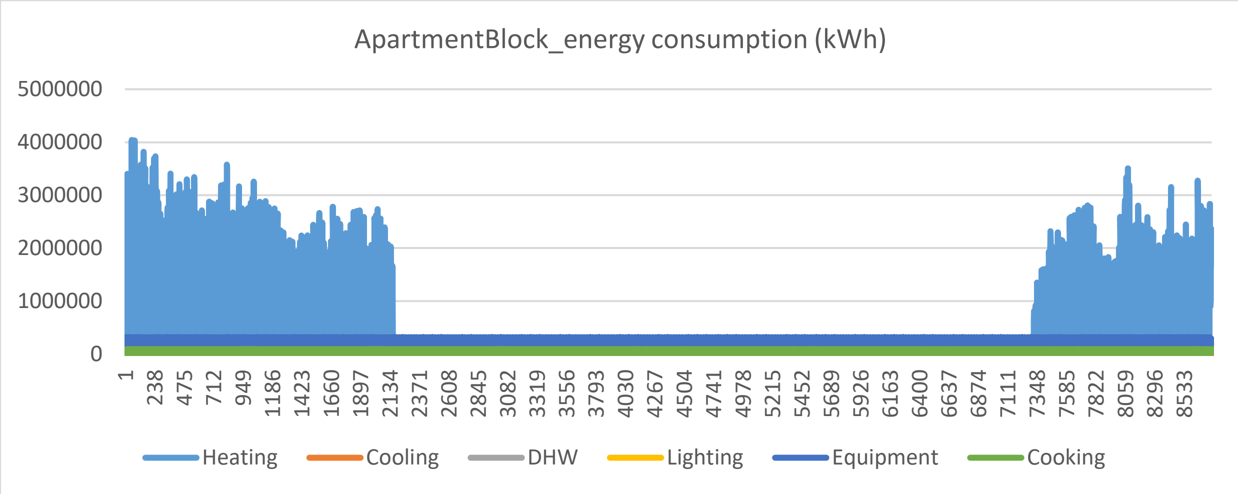

Figure 2. Visual example of the energy consumption results (hourly analysis of a year) for apartment block building category.

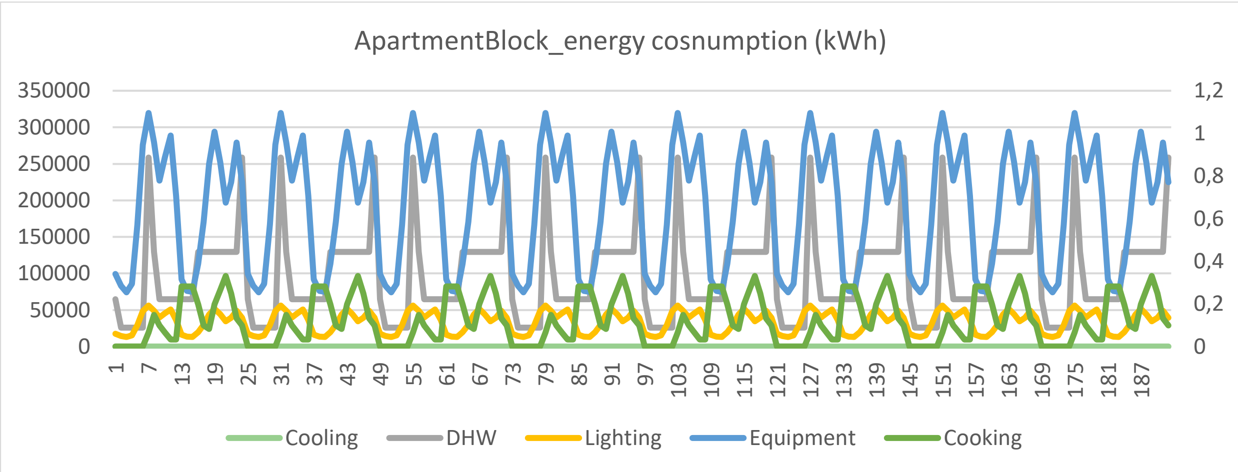

Figure 3. Visual example of the energy consumption results (hourly) for apartment block building category.

4. Implemented features

- The model allows building stock simulations for any region in Europe avoiding most of the building data collection, treatment, and preprocessing.
- The model allows simulations of a given year on an hourly basis (based on the heating degree hours method) for both NUTS Level 2 and NUTS level 3 contained in the region.
- The model has an internal database that provides the most relevant data to perform basic simulations for any region in Europe. It also offers the possibility of substituting these values for others proposed by the modeler in case more specific information is available.
- Through new simulations of future years, the model allows to evaluate the behavior of the building stock for different scenarios that consider different degrees of deployment of certain technologies.
- Regarding the coordination with the rest of the models of higher scales that contemplate the behavior of the building sector, the developed model maintains a coherence with them using a structure of disaggregation of building typologies, final uses and fuels used, as well as allowing the use of the outputs of the simulation models at higher scales to adjust some of the parameters of the regional model.
- The model offers a high degree of disaggregation for both the calculation and the presentation of results based on the following structure: Use - age - archetype - final use – fuel

Regarding the last point mentioned, the following is a more detailed description of the disaggregation that the model allows through the modeling structure used.

As for the disaggregation in the residential sector, the table below shows the degree of disaggregation included for each of the 49 categories included in the initial characterization phase.

Table1. Degree of disaggregation included for each of the 49 residential building categories

Sector	Use	Period	Archetype	Energy use	Share per Technology	Share of Fuels
Residential	Apartment Block	Pre-1945	% Archetype1 % Archetype2 % Archetype3	Heating Cooling DHW Cooking Lighting Equipment	% of each corresponding technology per energy use	% of each fuel corresponding to each technology and energy use

```
BUILDING_TYPOLOGY = {
  'Sector': 'Residential',
  'Use': 'Apartment Block',
  'Periods': ['Pre-1945'],
  'Archetypes': {'Archetype1': 0, 'Archetype2': 0, 'Archetype3': 0},
  'Energy_uses': {
    'Heating': {'Tech': {}, 'Fuels': {}},
    'Cooling': {'Tech': {}, 'Fuels': {}},
    'DHW': {'Tech': {}, 'Fuels': {}},
    'Cooking': {'Tech': {}, 'Fuels': {}},
    'Lighting': {'Tech': {}, 'Fuels': {}},
    'Equipment': {'Tech': {}, 'Fuels': {}}
  }
```

In the same way, for service sector the table below shows the degree of disaggregation included for each of the 49 categories included in the initial characterization phase.

Table2. Degree of disaggregation included for each of the 49 service building categories

Sector	Use	Period	Archetype	Energy use	Share per Technology	Share of Fuels
Service	office	Pre-1945	Archetype1	Heating Cooling DHW Cooking Lighting Equipment	% of each corresponding technology per energy use	% of each fuel corresponding to each technology and energy use

And finally, below it can be observed the complete disaggregation of building uses and ages considered by the model. It should be noted that each of these categories is further disaggregated by archetypes as shown in the tables above.

```
BUILDING_TYPOLOGY = {
  'Residential': {
    'Apartment Block': ['Pre-1945', '1945-1969', '1970-1979', '1980-1989', '1990-1999', '2000-2010', 'Post-2010'],
    'Single family- Terraced houses': ['Pre-1945', '1945-1969', '1970-1979', '1980-1989', '1990-1999', '2000-2010', 'Post-2010']
  },
  'Service': {
    'Other non-residential buildings': ['Pre-1945', '1945-1969', '1970-1979', '1980-1989', '1990-1999', '2000-2010', 'Post-2010'],
    'Offices': ['Pre-1945', '1945-1969', '1970-1979', '1980-1989', '1990-1999', '2000-2010', 'Post-2010'],
    'Hotels and Restaurants': ['Pre-1945', '1945-1969', '1970-1979', '1980-1989', '1990-1999', '2000-2010', 'Post-2010'],
    'Trade': ['Pre-1945', '1945-1969', '1970-1979', '1980-1989', '1990-1999', '2000-2010', 'Post-2010'],
    'Sport': ['Pre-1945', '1945-1969', '1970-1979', '1980-1989', '1990-1999', '2000-2010', 'Post-2010'],
    'Education': ['Pre-1945', '1945-1969', '1970-1979', '1980-1989', '1990-1999', '2000-2010', 'Post-2010'],
    'Health': ['Pre-1945', '1945-1969', '1970-1979', '1980-1989', '1990-1999', '2000-2010', 'Post-2010']
  }
}
```

5. Core assumptions

5.1. General assumptions

- Following the Energy Performance of Buildings Directive[1], static equations are used to determine the heating, cooling, and domestic hot water (DHW) energy demand. The methodology is based on the Degree-Days method. However, to obtain a more detailed analysis, the calculation is done on an hourly basis and considers internal gains, solar gains, ventilation losses[2].
- The model does not allow optimization. It is a simulation model that allows the evaluation of prospective exploratory scenarios based on the narratives set for each proposed scenario.
- Once the information is collected for each building, the model groups areas/geometries of buildings based on a clustering of buildings according to their form factor. Three archetypes are considered for each age range of residential buildings and a single representative archetype for each sub-use of the tertiary sector (7 subcategories of tertiary buildings are considered).
- Buildings of industrial use do not fall within the scope of the model. However, their geometric characteristics and geographic distribution across the region are considered to rule out areas available for large-scale solar technology use in the solar module.
- Limitations in terms of technologies or measures considered in the generation of future scenarios: Retrofitting measures are contemplated differentiating, facades, roofs, and windows (differentiating 3 levels; high, medium, low), replacement of heating and cooling systems, improvement of performances, installation of solar photovoltaic and thermal systems. To include measures related to district heating and cooling systems, they must be introduced as a new technology defining their share of fuel mix, performance, as well as other key parameters representative for the case study.

5.2. Mathematical description

The most relevant aspects of the mathematical description of the model are detailed below. It should be noted that this model corresponds to the adaptation of a previous model, so that once the basic methodological basis has been described, detailed descriptions of previous studies will be redirected.

This section is composed of different subsections, each of them focused on the description of the most relevant methodological and mathematical aspects of each of the large blocks described in the model design philosophy section.

[1] Directive 2010/31/EU of the European Parliament and of the Council of 19 May 2010 on the energy performance of buildings. Official Journal of the European Union 2010;L153:13–35.

[2] Oregi, X., Hermoso, N., Arrizabalaga, E., Mabe, L., Muñoz, I. (2018). Sensitivity assessment of a district energy assessment characterisation model based on cadastral data. *Energy Procedia*, 147, pp.181–188. Available at: https://doi.org/10.1016/j.egypro.2018.07.053

- Pre-processing of basic data of the building stock

a) BUILDINGS

Pre-processing of building-level information

This section describes the process followed for the preprocessing and treatment of the basic information of each building in the selected region. This is a process that the model executes without the need of inputs or user intervention. It is only necessary that the user selects the region to be evaluated.

The method is mainly based on the information available from OpenStreetMap[3]. This provides relevant information for the selected region regarding the geometry of the buildings and their use. However, this information has many gaps that are being filled in with other public data sources. Information available from the EU Building Stock Observatory[4] is used (which is the main classification used by the country level ESM tools to which the new model will be linked), as well as information from sources such as the Global Human Settlement Layer (GSH) data[5] and the Hotmaps project[6].

The method is integrated in a Python script capable of preparing the building stock energy data for any region of Europe. It provides relevant information per building of the region selected including the age, height, gross floor area, geometric shape and use and sub-use for residential and tertiary in a geolocated way.

Thus, when the user selects the NUTS Level 2 region to be evaluated, this process cuts the information corresponding to the geographical limits established for each of the information layers used. In a second phase, the information of each layer is processed to associate the characteristics to be extracted from it for one of the buildings contemplated in the region. This process can be replicated for an analysis to be performed at the NUTS level3 disaggregation level.

[3] OpenStreetMap data." https://download.geofabrik.de/europe.html (accessed Jun. 11, 2024).

[4] European Union, "EU Building Stock Observatory database." [Online]. Available: https://energy.ec.europa.eu/topics/energy-efficiency/energy-efficient-buildings/eu-building-stock-observatory_en.

[5] "GHSL - Global Human Settlement Layer." https://human-settlement.emergency.copernicus.eu/download.php (accessed Jun. 11, 2024)

[6] Hotmaps project." https://www.hotmaps.eu/map (accessed Jun. 11, 2024).

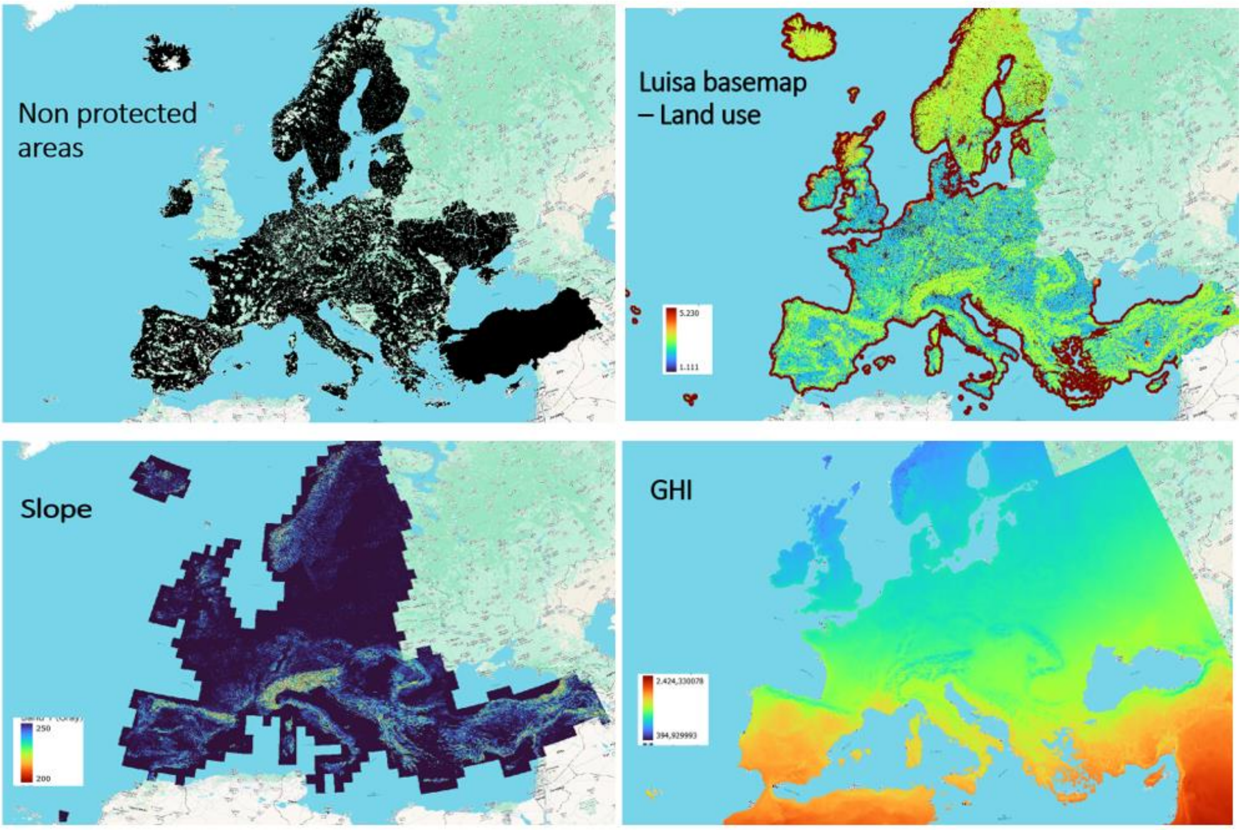

Figure 4. Visual example of different information layers that are used in different processes (buildings data preprocessing, building integrated SPV and SPV plants.

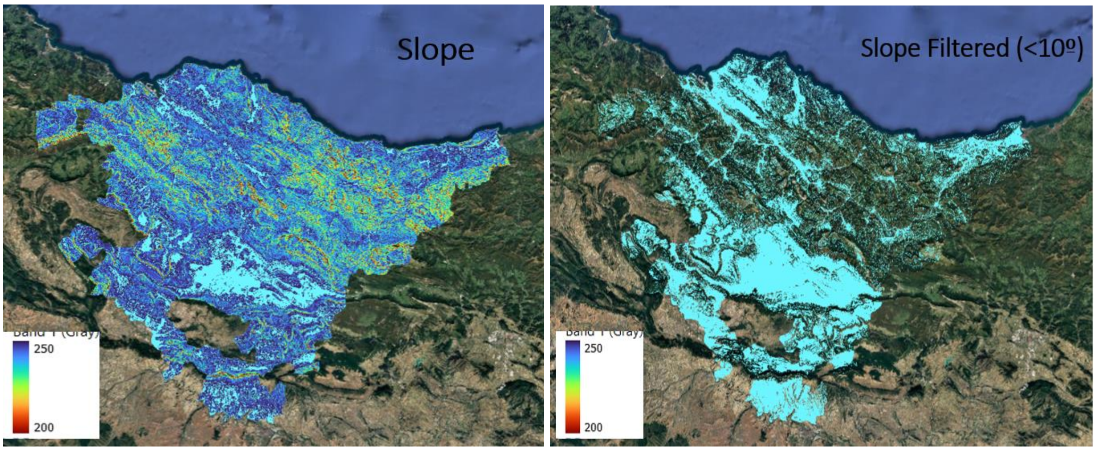

Figure 5. Visual example of one of the information layers (slope) before and after applying the necessary filters as part of the data processing.

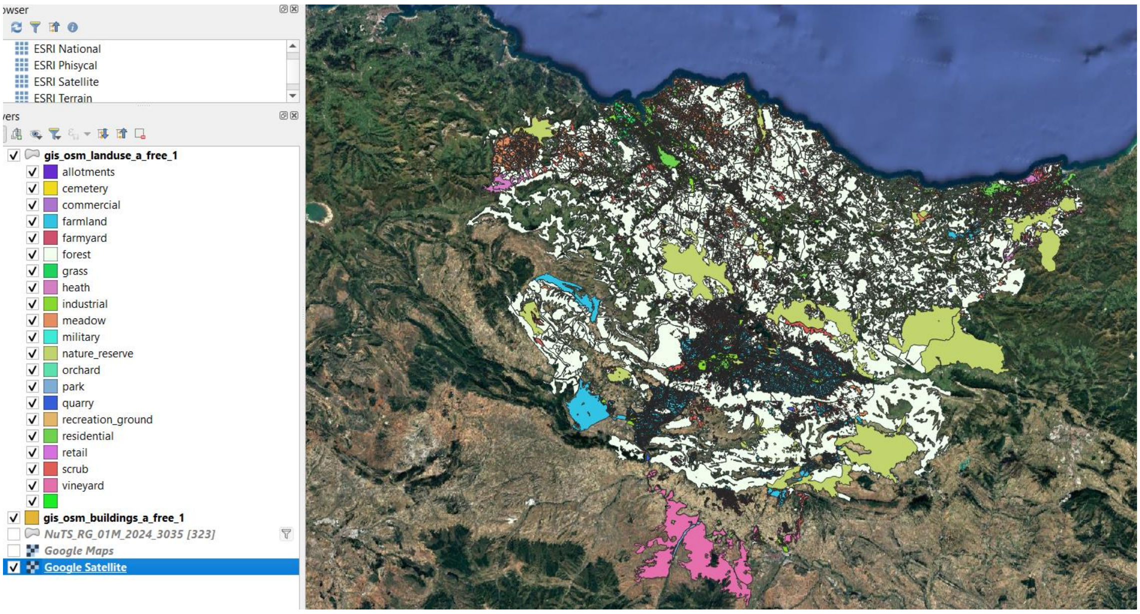

Figure 6. Visual example of the processing of the Geofabrik land use layer for a specific region.

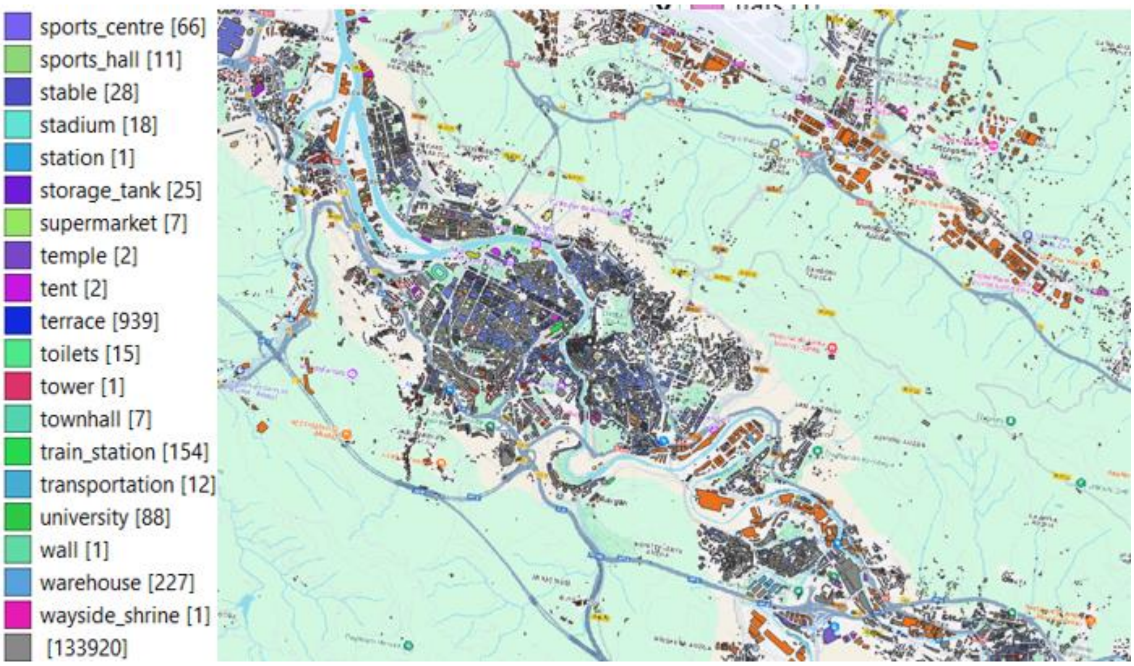

Figure 7. Visual example of the processing of the Geofabrik building use layer for a specific region/city.

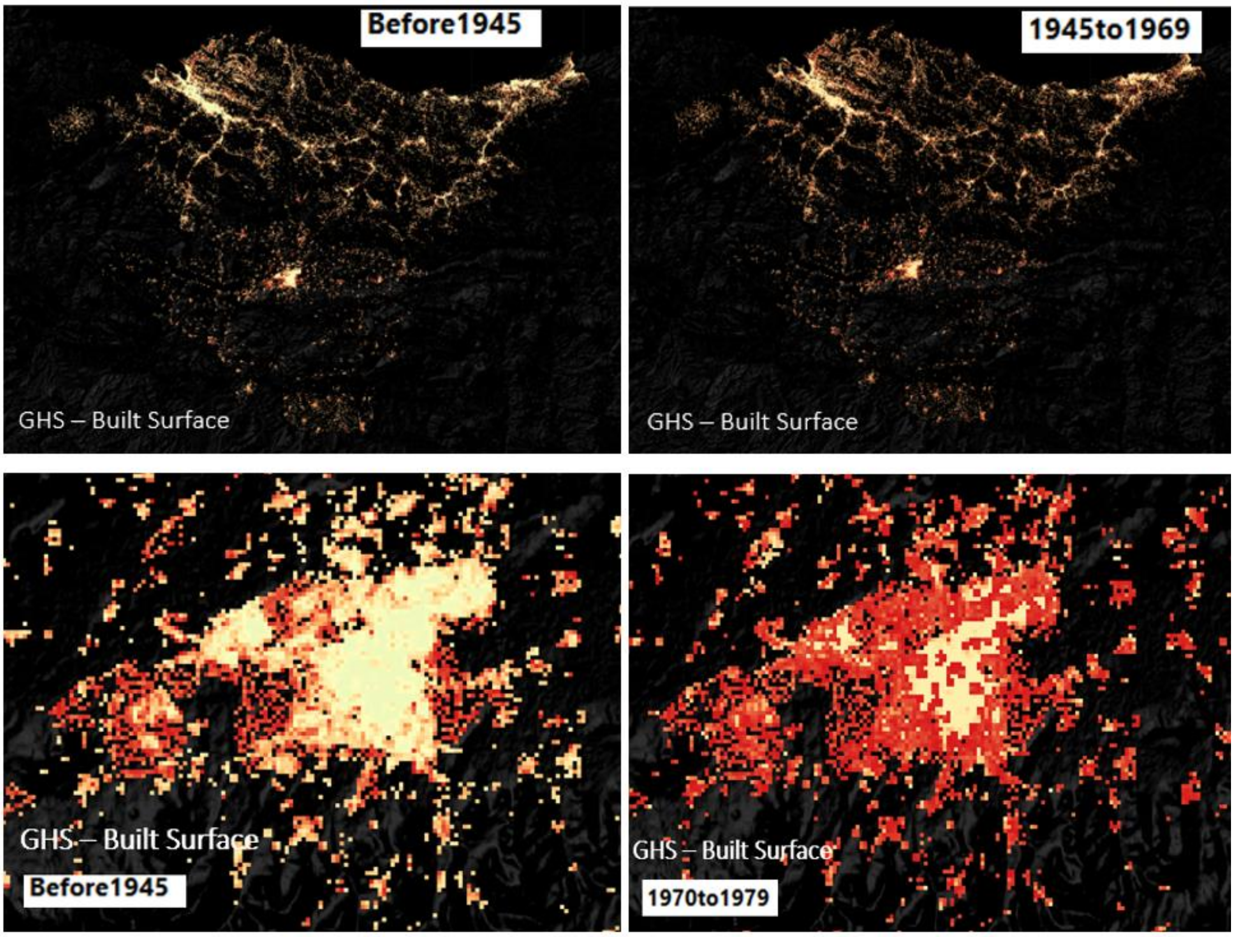

Figure 8. Visual example of the processing of the GHS – built surface layer for a specific region.

The performance of this development has been evaluated by comparing the results obtained with respect to the results obtained using a much more detailed and time-consuming approach which requires access to information based on cadastral data. This comparison has been carried out for the case of the city of Bilbao (a relevant city in the Basque Country, Spain), being aware that this is a very demanding comparative study as it evaluates the performance of the development for a city scale, which is more demanding than the regional scale. More details can be consulted in a paper dedicated to this section Data processing for georeferenced building characterization aimed at energy modelling of different NUTS level 2 regions of Europe[7].

Information processing for the generation of archetypes

In a next step building archetypes for each building typology according to their use and age are generated. In this way, through clustering methods, this process defines the number of building archetypes that are the most representative for each building typology (according to the use and age disaggregation). This

[7] E. Arrizabalaga[1], N. Hermoso[1], J. Pedrero[1], I. Muñoz[1], L. Mabe[1] (2024) Data processing for georeferenced building characterization aimed at energy modelling of different NUTS level 2 regions of Europe. European Climate and Energy Modelling Platform 2024 (ECEMP) Brussels and online 16th to 17th October 2024

simplifies the calculation of energy demands and consumption of the building stock regional energy model as buildings meeting certain characteristics can be grouped together to perform the energy calculation and not have to do it for each individual building. This is of greater interest as the scale of analysis is extended beyond the urban area.

Figure 9 shows the processing of this generated information to identify representative building archetypes which serve as additional input data to the building stock model, and which will contribute to accelerate energy demand calculations in the following steps of the energy planning process. For this purpose, the KMeans clustering algorithm[8] has been applied for each of the branches of the disaggregated tree of the building sector.

Once the clustering has been performed, the algorithm selects the building corresponding to the centroid of each cluster as the representative building to complete the final characterization of each archetype.

In this case, the resulting clusters for the case of building blocks in the residential sector by periods are shown in the figure below. The main parameters used for the grouping were the form factor, the ratio of the external facade area to the total area of the façade (including adjacent facades), and the height. The building that corresponds to the centroid of each cluster is the one that will serve as the basis for defining the characteristics of the archetype used for the calculation.

[8] https://scikit-learn.org/1.5/modules/clustering.html#k-means

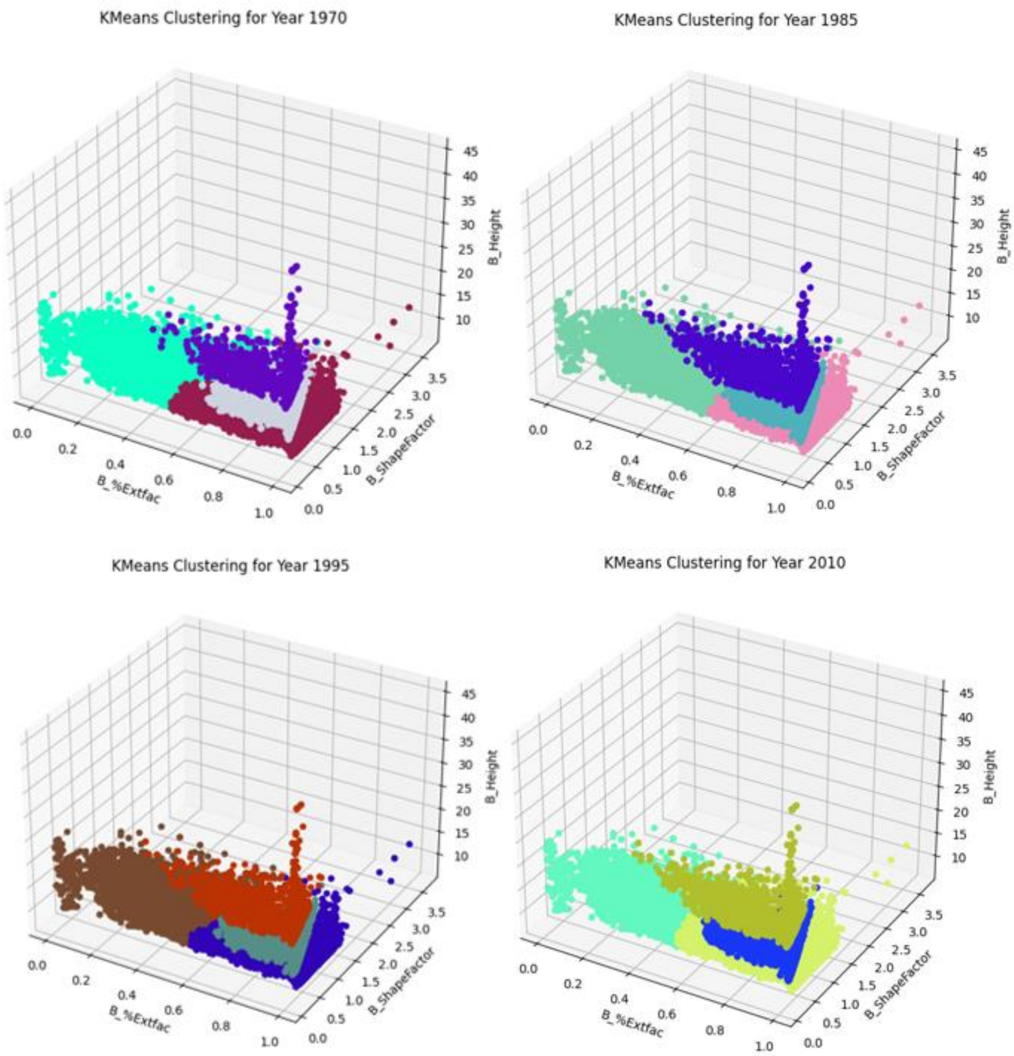

Figure 9. Visual example of the clusters for the residential sector by building periods for the city of Bilbao.

b) SOLAR

This section describes the information preprocessing process used as input to the solar building technology assessment module.

- Pre-processing of input data used for the solar process.

The process determines for each selected region which buildings are suitable for the potential implementation of solar technology.

For each region a raster file is generated with the buildings of potential radiation differentiated by different radiation ranges or thresholds. This is intended to obtain more criteria to determine which are the most interesting areas in which to start deploying solar technology once the alternative scenarios are being generate.

As output of this process, the results obtained in terms of the region ID (NUTS Level 3 and NUTS Leve 2), the coordinates of the region centroid, the total area, the maximum radiation, the average radiation, and the area corresponding to each of the buildings corresponding to each radiation threshold measured above are stored.

This information is used in later phases to calculate solar generation, taking as inputs data with a high degree of disaggregation.

The different layers of information used in the process described above are listed in more detail below.

- Global solar radiation atlas GHI at 200x200m resolution[9]
- Land use map: 100x100m. Selected from this layer the uses corresponding to each building typology.

The following is a more detailed description of the preprocessing process according to its logic and processing sequence.

Data Preparation and Clipping

- Clip the NUTS (Nomenclature of Territorial Units for Statistics) regions layer to the selected NUTS 2 or NUTS 3 region.
- Clip the following layers to the selected region: (Land use, Solar radiation)

Land Use Layer Filtering

The land use layer is filtered to identify specific land cover types suitable for solar installation:

- User-defined codes for suitable building types (e.g., Single family houses, Apartments, Health, Hotels, Offices, Sport, Trade) are used to filter the layer.
- Pixels corresponding to the specified land use codes are set to 1, while all others are set to 0.

Raster Alignment and Standardization

All layers are aligned and adjusted to ensure consistent resolution, projection system, and size across datasets.

Regional Analysis and Reporting

The code performs a regional analysis based on the potential radiation areas:

- Radiation thresholds are defined from a starting value to the maximum raster value, incrementing by 100.
- A Data Frame is created to store results, including region ID, centroid coordinates, total area, maximum radiation, average radiation, and area corresponding to each radiation threshold.
- The code iterates over each NUTS 3 region, calculating values for each radiation threshold.
- Results are saved as a CSV file in the same directory as the input raster file.

Region,Centroid_X,Centroid_Y,Total_Area,Max_Radiation,Average_Radiation,Threshold,Area_m2,Median_Radiation,Median_Radiation_X,Median_Radiation_Y

[9] https://globalsolaratlas.info/download/world

- ### Calculation of energy needs and energy consumption of buildings in the region

Buildings

The most relevant mathematical equations used for the calculation of the energy needs of buildings are described below. These algorithms are applicable to different building typologies, from residential buildings to tertiary buildings.

It should be noted however that for each case, the input parameters considered in this building stock energy model are different depending on the use and age of the buildings.

$$AHD_k = \sum_{i,j=1}^{8760} \left(HDH_{i,j} \times A_k \times U_k - Gains_{i,j} + hventilation\ losses_{i,j} \times (1 - n_{HR}) \right) \cdot heating\ schedule_{i,j} \qquad \text{(Eq. 1}$$

Where AHD_k is the annual heating useful energy demand (kWh/year), $HDH_{i,j}$ is the heating degree hours (°C), A_k is the envelope element surface (m^2), U_k is the thermal transmittance ($W/(m^2 \cdot K)$), ηHR is the heat recovery system efficiency (%), i is the hour of the day and j is the day of the year.

A similar procedure is used for the calculation of the annual cooling demand of buildings but in this case the heating degree hours are replaced by the cooling degree hours, the heating ventilation losses are replaced by the cooling ventilation losses, and the heating schedule by the cooling schedule.

$$ACD_k = \sum_{i,j=1}^{8760} \left(CDH_{i,j} \times A_k \times U_k + Gains_{i,j} + cventilation\ losses_{i,j} \times (1 - n_{HR}) \right) \cdot cooling\ schedule_{i,j} \qquad \text{(Eq. 2)}$$

Where ACD_k is the annual cooling useful energy demand (kWh/year), $CDH_{i,j}$ is the cooling degree hours (°C), A_k is the envelope element surface (m2), U_k is the thermal transmittance ($W/(m^2 \cdot K)$), ηHR is the heat recovery system efficiency (%), i is the hour of the day and j is the day of the year.

Finally, the annual domestic hot water demand is determined by multiplying the annual DWH demand per square meter, the gross floor area of the building and the normalized usage factor of the DHW.

$$DHWD_k = \sum_{i,j=1}^{8760} DHW\ demand_k \times NHA_k \times \frac{Hourly\ usage\ factor_{DHW_{i,j}}}{\sum_{i,j=1}^{8760} Hourly\ usage\ factor_{DHW_{i,j}}} \qquad \text{(Eq. 3)}$$

Where $DHWD_k$ is the annual domestic hot water useful energy demand (kWh/year), DHW_k is the domestic hot water demand (kWh/m^2), NHA_k is the net heated area (m^2), i is the hour of the day and j is the day of the year.

The values of the parameters used in the equations vary according to the location, age, or use of the building, as shown in the table 1.

Table3. Dependence of the parameters according to the characteristics of the buildings

	Schedules	Internal gains	WWR	U-value	Ventilation losses	Solar gains	DHW demand
Location				X	X	X	
Age				X	X		
Use	X	X	X	X		X	X

Internal database

The building stock model incorporates a database with the most relevant information that is necessary for its implementation. It combines information downloaded from different sources such as Demand.ninja[10] for weather (hourly data), the EU Buildings Stock Observatory[11] for basic data such as thermal transmittances by building type, age, and location, or the European Commission, Joint Research Centre (JRC) Dataset[12] to obtain preliminary shares of technology use by final energy use for each building typology and energy performances of technologies for energy consumption calculation.

Table4. Structure of technology use by final energy use for each building typology used to define the share of technologies as preliminary data of the model per country.

ES - Final energy consumption per surface area	2021	Code
Final energy consumption (kWh / sqm)	61,2	FEC_per_sqm.kWh.ES.Res.HH.Thermal
Space heating	41,6	FEC_per_sqm.kWh.ES.Res.HH.Thermal.SH
Solids	41,0	FEC_per_sqm.kWh.ES.Res.HH.Thermal.SH.Solids.Solids
Liquified petroleum gas (LPG)	47,7	FEC_per_sqm.kWh.ES.Res.HH.Thermal.SH.LPG.LPG
Diesel oil	47,1	FEC_per_sqm.kWh.ES.Res.HH.Thermal.SH.Oil.Oil_LiqBio
Natural gas	41,9	FEC_per_sqm.kWh.ES.Res.HH.Thermal.SH.Gas.NG_Biogas
Biomass	45,5	FEC_per_sqm.kWh.ES.Res.HH.Thermal.SH.Biomass.Biomass_Waste
Geothermal	36,8	FEC_per_sqm.kWh.ES.Res.HH.Thermal.SH.Geo.Geo
Distributed heat	-	FEC_per_sqm.kWh.ES.Res.HH.Thermal.SH.DistrHeat.Steam_Distr
Advanced electric heating	9,1	FEC_per_sqm.kWh.ES.Res.HH.Thermal.SH.AdvElc.Elec
Conventional electric heating	35,2	FEC_per_sqm.kWh.ES.Res.HH.Thermal.SH.ConvElc.Elec
Electricity in circulation	0,4	FEC_per_sqm.kWh.ES.Res.HH.Thermal.SH.TotalCirculation.Elec
Space cooling	0,9	FEC_per_sqm.kWh.ES.Res.HH.Thermal.SC
Air conditioning	0,9	FEC_per_sqm.kWh.ES.Res.HH.Thermal.SC.AC.Elec
Water heating	12,2	FEC_per_sqm.kWh.ES.Res.HH.Thermal.WH
Solids	12,0	FEC_per_sqm.kWh.ES.Res.HH.Thermal.WH.Solids.Solids
Liquified petroleum gas (LPG)	12,3	FEC_per_sqm.kWh.ES.Res.HH.Thermal.WH.LPG.LPG
Diesel oil	13,8	FEC_per_sqm.kWh.ES.Res.HH.Thermal.WH.Oil.Oil_LiqBio
Natural gas	13,2	FEC_per_sqm.kWh.ES.Res.HH.Thermal.WH.Gas.NG_Biogas
Biomass	12,2	FEC_per_sqm.kWh.ES.Res.HH.Thermal.WH.Biomass.Biomass_Waste
Geothermal	11,3	FEC_per_sqm.kWh.ES.Res.HH.Thermal.WH.Geo.Geo
Distributed heat	-	FEC_per_sqm.kWh.ES.Res.HH.Thermal.WH.DistrHeat.Steam_Distr
Electricity	6,9	FEC_per_sqm.kWh.ES.Res.HH.Thermal.WH.Elc.Elec
Solar	8,5	FEC_per_sqm.kWh.ES.Res.HH.Thermal.WH.TotalSolar.Solar
Cooking	6,6	FEC_per_sqm.kWh.ES.Res.HH.Thermal.CO
Solids	7,4	FEC_per_sqm.kWh.ES.Res.HH.Thermal.CO.Solids.Solids
Liquified petroleum gas (LPG)	8,0	FEC_per_sqm.kWh.ES.Res.HH.Thermal.CO.LPG.LPG
Natural gas	8,1	FEC_per_sqm.kWh.ES.Res.HH.Thermal.CO.Gas.NG_Biogas
Biomass	8,1	FEC_per_sqm.kWh.ES.Res.HH.Thermal.CO.Biomass.Biomass_Waste
Electricity	5,6	FEC_per_sqm.kWh.ES.Res.HH.Thermal.CO.Elc.Elec

10 https://renewables.ninja/

11 European Commission. EU Buildings Stock Observatory. Available: http://ec.europa.eu/energy/en/eubuildings

12 Rozsai, Mate; Jaxa-Rozen, Marc; Salvucci, Raffaele; Sikora, Przemyslaw; Tattini, Jacopo; Neuwahl, Frederik (2024): JRC-IDEES-2021. European Commission, Joint Research Centre (JRC) [Dataset] PID: http://data.europa.eu/89h/82322924-506a-4c9a-8532-2bdd30d69bf5

It should be noted that all these parameters together with the use profiles of each building typology are susceptible to being substituted by more precise input data that the modeler may have available for the region being evaluated. As a conceptual example, the simplified form of data entry for each building typology is shown for each building typology in terms of weekday consumption profiles by final energy use.

Table5. Weekday consumption profiles by final energy use.

Weekday	Heating	Cooling	Lighting	Equipment	Occupancy	DHW	Cooking
0:00	0,0	0,2	0,3	0,3	1,0	0,3	0,0
1:00	0,0	0,1	0,3	0,3	1,0	0,1	0,0
2:00	0,0	0,1	0,2	0,2	1,0	0,1	0,0
3:00	0,1	0,1	0,3	0,3	1,0	0,1	0,0
4:00	0,0	0,2	0,5	0,5	1,0	0,1	0,0
5:00	0,1	0,6	0,9	0,9	1,0	0,1	0,0
6:00	0,9	0,9	1,0	1,0	1,0	1,0	0,2
7:00	1,0	1,0	0,9	0,9	0,5	0,5	0,5
8:00	0,9	0,9	0,7	0,7	0,3	0,3	0,3
9:00	0,5	0,7	0,8	0,8	0,3	0,3	0,2
10:00	0,6	0,4	0,9	0,9	0,3	0,3	0,1
11:00	0,7	0,2	0,6	0,6	0,3	0,3	0,1
12:00	0,6	0,2	0,3	0,3	0,3	0,3	0,9
13:00	0,7	0,1	0,2	0,2	0,3	0,3	0,9
14:00	0,7	0,1	0,2	0,2	0,3	0,3	0,9
15:00	0,7	0,1	0,4	0,4	0,5	0,5	0,6
16:00	0,7	0,2	0,5	0,5	0,5	0,5	0,3
17:00	0,7	0,5	0,8	0,8	0,5	0,5	0,3
18:00	0,8	0,9	0,9	0,9	0,5	0,5	0,6
19:00	1,0	1,0	0,8	0,8	0,5	0,5	0,8
20:00	0,9	0,9	0,6	0,6	0,5	0,5	1,0
21:00	0,9	0,7	0,7	0,7	0,5	0,5	0,8
22:00	0,7	0,4	0,9	0,9	0,5	0,5	0,4
23:00	0,5	0,2	0,7	0,7	1,0	1,0	0,3

Building integrated solar systems

Once potential areas of each building type in each NUTS3 region are categorized by intervals of 100W/m^2 of Global Horizontal Irradiance (GHI), those with higher GHI are selected until reaching the area, power capacity or investment required by the user at the input.

The required power capacity is calculated from the investment considering the following unit costs. These can be also configured by the user.

Table 6. Unit cost in €/Wp for each building typology.

Category	Subcategory	Unit cost (€/Wp)
Residential	Single-family house	1.2
	Apartment	0.8
Commercial	Health	0.8
	Hotels	0.8
	Offices	0.8
	Sports	0.8
	Trade	0.8

The required area is calculated from the power capacity considering the following system efficiencies and use factors, understood as the rate of area in the building to install PV. These can be also configured by the user.

Table 7.. System efficiency and use factor in % for each building typology.

Category	Subcategory	System efficiency (%)	Use factor (%)
Residential	Single-family house	25	60
	Apartment	15	80
Commercial	Health	15	80
	Hotels	15	80
	Offices	15	80
	Sports	15	80
	Trade	15	80

For the location corresponding to the median of GHI of each selected area of each NUTS3 the Photovoltaic (PV) generation profile is estimated and then aggregated for all the selected areas of all the NUTS3 to compute the PV generation profile in buildings for the whole NUTS2 region.

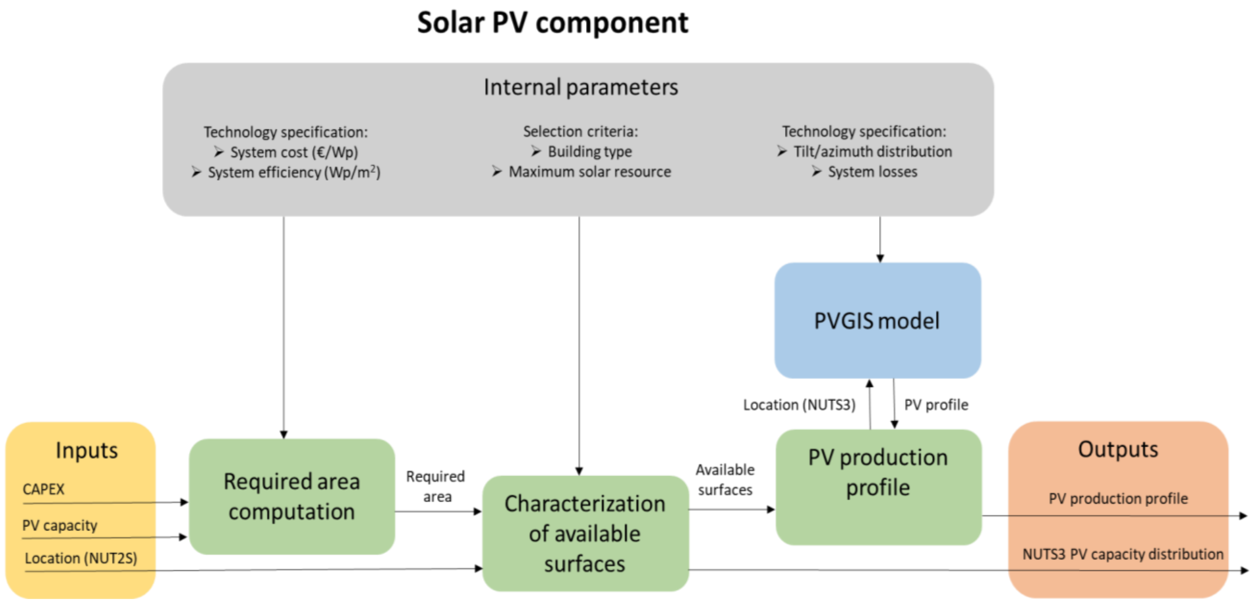

Figure 10. General scheme of Solar PV component in buildings.

The tool used to estimate the Photovoltaic generation in buildings is PVGIS[13], a free web application that allows the user to get data on solar radiation and photovoltaic system energy production. This is automatedly accessible via API.

Firs of all, PVGIS makes use of reanalysis-based solar radiation data sets to estimate the solar radiation arriving at the earth surface. Data. Reanalysis data are calculated using numerical weather forecast models, re-running the models for the past and making corrections using the known meteorological measurements. The output of the models is a large number of meteorological quantities, often including the solar irradiance at ground level. More concretely, ECMWF ERA-5, produced by the European Centre for Medium-range Weather Forecast (ECMWF), has global coverage at a resolution of about 30km, and includes both global and direct solar irradiance. At the time of writing, only the time period 2005-2020 has been in PVGIS along with global coverage.

The satellite-based calculation described above produces values of global and beam irradiance on a horizontal plane. However, modules and PV systems are generally installed at an inclined angle regarding the horizontal plane or on tracking systems, so as to maximize the received in-plane irradiance. Therefore, the satellite retrieved irradiance values are not representative of the solar radiation available at the module surface, and it becomes necessary to estimate the in-plane irradiance.

There are several models in the scientific bibliography which use as input data the irradiance values on the horizontal plane of global and diffuse and/or beam irradiance components, to estimate the values of the beam and diffuse components on tilted surfaces. A comparison of some of these models can be found in *Gracia Amillo and Huld, 2013*. The estimation model implemented in PVGIS is the one developed by Muneer T. (1990) *,* which can be classified as anisotropic of two components.

In the case of buildings, fixed mounting structure has been considered with optimal azimuth (South oriented) and a tilt of 20º as a trade-off between energy yield and system efficiency. Only in the case of single-family houses 20% of the systems are East-oriented and other 20% West-oriented, as these are normally adapted to roof orientation.

On the other hand, PVGIS uses information about the elevation of the terrain with a resolution of 3 arc-seconds (about 90m). This means that for every 90m we have a value for the ground elevation. From these data we have calculated the height of the horizon around each geographical location. These data are then used to calculate the times when the sun is shadowed by hills or mountains. When this happens, the solar radiation is then calculated using only the diffuse part of the radiation. Please notice that with a resolution of ~90m the calculations in PVGIS cannot consider the effects of shadows from nearby objects such as houses or trees. Therefore, a higher “system loss” than usual must be considered for PV in buildings.

Indeed, once the amount of solar radiation that arrives at the PV modules is computed, the different effects that influence PV output and how they are calculated in PVGIS.

- Shallow-angle reflection. This is calculated using a mathematical model described in (*Martin&Ruiz, 2001, Martin&Ruiz, 2013*). Generally, this effect causes a loss of 2-4% of the sunlight, though this will be lower for sun-tracking PV systems (*Huld et al., 2015*).
- Effects of changes in the solar spectrum. PVGIS has used solar radiation data from satellite that have been calculated for different spectral bands (*Mueller et al., 2012*) to calculate the effect of spectrum changes on the PV energy output.

[13]PVGIS data sources & calculation methods - European Commission

- PV power dependence on irradiance and module temperature. PVGIS calculates the effects of irradiance and module temperature using a model described in (*Huld et al., 2011*). Module temperature is treated in PVGIS using a model suggested by Faiman (*Faiman, 2008*) and the BIPV/BAPV configuration is selected in this case.
- System losses and degradation. Considering the system losses and the losses due to ageing PVGIS recommends a value of 14% for the "system loss" that the user gives as input to PVGIS. However, there are a number of other effects that can influence the energy output of PV systems. These effects are not included in the PVGIS calculations. Among these are: snow, soling, partial shadowing. Considering building environment, more prone to the appearance of these issues, a value of 20% is set for the "system loss" in this case.

Energy scenario configuration

For the configuration of future scenarios, the model allows a simulation of a given fixed year, allowing to configure the main characteristics of that year.

In the buildings sector, the model allows to consider the effect of the evolution of HDD and CDD in the future, as well as the construction of new buildings. Based on this configuration, the model makes it possible to determine the degree of deployment of different technologies or measures applicable to each typology and archetype of buildings defined.

Thus, for example, in the case of building retrofitting, the following figure serves to understand the logic behind the configuration of retrofitting scenarios for each building typology. And depending on the selected degree of retrofitting, the model adapts the thermal transmittance values of the building envelope to recalculate the new consumptions.

	Ref. level	Pre-1945	1945-1969	1970-1979	1980-1989	1990-1999	2000-2010	Post-2010
Apartment block	Low	85%	25%	13%	0%	0%	0%	0%
Single family- Terraced houses	High	50%	17%	0%	0%	0%	0%	0%
Offices	Medium	5%	5%	5%	5%	5%	0%	0%
Education	Medium	10%	10%	10%	10%	10%	10%	10%
Health	Low	0%	0%	0%	0%	0%	0%	0%
Trade	Low	0%	0%	0%	0%	0%	0%	0%
Hotels and Restaurants	Medium	0%	0%	0%	0%	0%	0%	0%
Other non-residential buildings	Medium	0%	0%	0%	0%	0%	0%	0%
Sport	Medium	0%	0%	0%	0%	0%	0%	0%

Figure 11. Logic behind the configuration of retrofitting scenarios for each building typology.

```
# Input parameters

# Definition of building types
BUILDING_TYPES = [
    'Apartment block',
    'Single family- Terraced houses',
    'Offices',
    'Education',
    'Health',
    'Trade',
    'Hotels and Restaurants',
    'Other non-residential buildings',
```

```
    'Sport'
]

# Construction periods
CONSTRUCTION_PERIODS = [
    'Pre-1945',
    '1945-1969',
    '1970-1979',
    '1980-1989',
    '1990-1999',
    '2000-2010',
    'Post-2010'
]

# Reference levels for building refurbishment
REFERENCE_LEVELS = {
    'Low',
    'Medium'
    'High'
}

# Percentage of surface area to be refurbished by building type and construction period
RENOVATION_PERCENTAGE = {
    'Apartment block': {
        'Pre-1945': 100,
        '1945-1969': 0,
        '1970-1979': 0,
        '1980-1989': 0,
        '1990-1999': 0,
        '2000-2010': 0,
        'Post-2010': 0
    },
    'Single family- Terraced houses': {
        'Pre-1945': 0,
        '1945-1969': 0,
        '1970-1979': 0,
        '1980-1989': 0,
        '1990-1999': 0,
        '2000-2010': 0,
        'Post-2010': 0
    },
    # ... (repeated for other building types)
}
```

Similarly, the configuration of technology substitution scenarios can be configured for each building typology according to the logic shown in the following table.

Thus, the described % is associated with a given built-up area, a given number of buildings, a given number of technologies replaced (number or installed capacity) and a given investment.

Table 8. New technology share configuration by end use for a typical building.

Energy service	Baseline configuration	Scenario configuration
_Space heating	100,00%	100,00%
Space heating_Solids	0,18%	Insert
Space heating_Liquified petroleum gas (LPG)	0,45%	Insert
Space heating_Diesel oil	12,62%	Insert
Space heating_Gas heat pumps		Insert
Space heating_Natural gas	23,53%	Insert
Space heating_Biomass	30,95%	Insert
Space heating_Geothermal	0,00%	Insert
Space heating_Distributed heat	14,66%	Insert
Space heating_Advanced electric heating	9,27%	Insert
Space heating_Conventional electric heating	7,32%	Insert
Space heating_BioOil	0,00%	Insert
Space heating_BioGas	0,00%	Insert
Space heating_Hydrogen	0,00%	Insert
_Electricity in circulation	1,03%	Insert
_Space cooling	7,80%	7,80%
Space cooling_Gas heat pumps	0,00%	Insert
Space cooling_Electric space cooling	100,00%	Insert
_Water heating	100,00%	100,00%
Water heating_Solids	0,05%	Insert
Water heating_Liquified petroleum gas (LPG)	0,31%	Insert
Water heating_Diesel oil	8,19%	Insert
Water heating_Natural gas	17,64%	Insert
Water heating_Biomass	17,69%	Insert
Water heating_Geothermal	0,00%	Insert
Water heating_Distributed heat	14,00%	Insert
Water heating_Electricity	28,40%	Insert
Water heating_BioOil	0,00%	Insert
Water heating_BioGas	0,00%	Insert
Water heating_Hydrogen	0,00%	Insert
Water heating_Solar	13,72%	Insert
_Cooking	100,00%	100,00%
Cooking_Solids	0,02%	Insert
Cooking_Liquified petroleum gas (LPG)	0,19%	Insert
Cooking_Natural gas	4,06%	Insert
Cooking_Biomass	2,85%	Insert
Cooking_Electricity	92,89%	Insert
_Appliances	100,00%	100,00%
Appliances_Electricity	1,0000	Insert
_Lighting	100,00%	100,00%
Lighting_Electricity	1,0000	Insert

iDesignRES

iDesignRES

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.