FlexPlan

Advanced methodology and tools taking advantage of storage and FLEXibility in transmission and distribution grid PLANning

Data set and planning criteria for the regional studies

D5.1

Distribution Level	PU
Responsible Partner	RSE
Checked by WP leader	Nuno Amaro (R&D Nester) – WP5 Leader Date: 15.09.2021
Verified by the appointed Reviewers	Iver Bakken Sperstad (SINTEF) – Date: 10.08.2021 Dragana Orlic (EKC) – Date: 02.08.2021
Approved by Project Coordinator	Gianluigi Migliavacca (RSE) Date: 15.09.2021

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 863819

Issue Record

Planned delivery date	30.09.2021
Actual date of delivery	15.09.2021
Status and version	FINAL

Version	Date	Author(s)	Notes
01	16.07.2021	Izabella Faifer (RSE) Nuno Amaro (R&D Nester) Raúl Rodríguez (TECNALIA) Babak Bardideh (KU LEUVEN) Dennis Schmid (Tu-Dortmund) Marco Rossi (RSE) Boris Brdjanin (EKC) Iver Bakken Sperstad (SINTEF) Michele Garau (SINTEF) Vegard Viken Kallset (SINTEF) Merkebu Zenebe Degefa (SINTEF) Matteo Paolo Costa (RSE) Benedetta Marmiroli (RSE) Alberte Bouso (N-SIDE)	Merged updates from RSE, R&D Nester, formatting and review
02	16.08.2021	Izabella Faifer (RSE) Babak Bardideh (KU LEUVEN) Dennis Schmid (Tu-Dortmund) Marco Rossi (RSE) Boris Brdjanin (EKC) Alberte Bouso (N-SIDE) Matteo Paolo Costa (RSE)	Updates from reviewers

		Nuno Amaro (R&D Nester)	

About FlexPlan

The FlexPlan project aims at establishing a new grid planning methodology considering the opportunity to introduce new storage and flexibility resources in electricity transmission and distribution grids as an alternative to building new grid elements. This is in line with the goals and principles of the new EC package *Clean Energy for all Europeans,* which emphasizes the potential usage of flexibility sources in the phases of grid planning and operation as alternative to grid expansion. In sight of this, FlexPlan creates a new innovative grid planning tool whose ambition is to go beyond the state of the art of planning methodologies, by including the following innovative features: integrated T&D planning, full inclusion of environmental analysis, probabilistic contingency methodologies replacing the N-1 criterion as well as optimal planning decision over several decades. However, FlexPlan is not limited to building a new tool but it also uses it to analyse six regional cases covering nearly the whole European continent, aimed at demonstrating the application of the tool on real scenarios as well as at casting a view on grid planning in Europe till 2050. In this way, the FlexPlan project tries to answer the question of which role flexibility could play and how its usage can contribute to reduce planning investments yet maintaining (at least) the current system security levels. The project ends up formulating guidelines for regulators and for the planning offices of TSOs and DSOs. The consortium includes three European TSOs, one of the most important European DSO group, several R&D companies and universities from 8 European Countries (among which the Italian RSE acting as project coordinator) and N-SIDE, the developer of the European market coupling platform EUPHEMIA.

Partners

Table of Contents

About FlexPlan 4

Table of Contents 5

List of Abbreviations and Acronyms 7

Executive Summary 8

1 Introduction 11

2 Applied planning and optimization criteria 13

3 Network data and modelling 15

3.1 Transmission and sub-transmission network 15

3.1.1 Iberian Peninsula 15

3.1.2 France and Benelux 18

3.1.3 Germany, Switzerland and Austria 21

3.1.4 Italy 22

3.1.5 Balkan region 22

3.1.6 Nordic region 23

3.2 Distribution network 25

3.2.1 Iberian Peninsula 25

3.2.2 France and Benelux 26

3.2.3 Germany, Switzerland and Austria 26

3.2.4 Italy 27

3.2.5 Balkan region 28

3.2.6 Nordic region 29

4 Generation and load input data and its localization 30

4.1 Generation 30

4.1.1 Iberian Peninsula 30

4.1.2 France and Benelux 33

4.1.3 Germany, Switzerland and Austria 33

4.1.4 Italy 34

4.1.5 Balkan region 35

4.1.6 Nordic region 36

4.2 Load 37

4.2.1 Iberian Peninsula 37

4.2.2 France and Benelux 38

4.2.3 Germany, Switzerland and Austria 39

4.2.4 Italy 39

4.2.5 Balkan region 39

4.2.6 Nordic region 40

5 Environmental indices and input data 41

5.1 Carbon footprint data ... 42
5.2 Air quality ... 43
5.3 Environmental models ... 43
6 Costs and economic data ... 46
7 Common post processing of the market results ... 48
7.1 Used approach for adaptation of Pan-European market results ... 48
8 Common pre-processing procedures for the case studies ... 51
8.1 ICT aspects and JSON file structure ... 51
8.2 JSON file structure ... 52
8.3 Used approach for creation network models in JSON format ... 53
9 Conclusions ... 55
10 Annex 1 – Network modelling assumptions summary ... 56
11 Annex 2 – Environmental data template ... 58
12 References ... 59

List of Abbreviations and Acronyms

Abbreviation/Acronym	Meaning
AC	Alternating Current
API	Application Programming Interface
CAMx	Comprehensive Air Quality model with extensions
CBA	Cost-benefit Analysis
CGMES	Common Grid Model Exchange Standard
CTM	Chemistry and transport model
DC	Direct Current
DDM	Decoupled Direct Method for sensitivity analysis in a three-dimensional air quality model
DSO	Distribution System Operator
EHV	Extra High Voltage
ENTSO-E	European Network of Transmission System Operators for Electricity
EMAS	European Eco-management and Audit Scheme
EMEP/EEA	Air pollutant emission inventory guidebook
ECMWF	European Centre for Medium-Range Weather Forecasts
EV	Electric vehicle
HV	High voltage
IPCC	Intergovernmental Panel on Climate Change
JSON	JavaScript Object Notation, data format
LCP	Large combustion plants
MCA	Multi-criteria Analysis
MILES	Model of International Energy Systems
OHL	Overhead line
OPF	Optimal Power Flow
OSM	OpenStreetMap
PV	Photovoltaics
PRTR	Pollutant Release and Transfer Register
RES	Renewable Energy Source
TSO	Transmission System Operator
TYNDP	10-year network development plan
UGC	Underground Power Cables

Executive Summary

This deliverable includes the results of the preparatory work for data collection and re-elaboration in order to set up the six regional case studies to be carried out in the FlexPlan project.

The present document provides an overview of the data set and planning criteria that will be used for the regional studies. This includes data of transmission and distribution networks for 2025 (as a starting point) and an adaptation of the market results for 2030, 2040 and 2050 that were generated as output of the FlexPlan pan-European studies [1][2]. The resulting data set will initialize a planning process starting from non-expanded OPF simulations (network without reinforcements or new flexibility devices), as presented in the following figure 1-1.

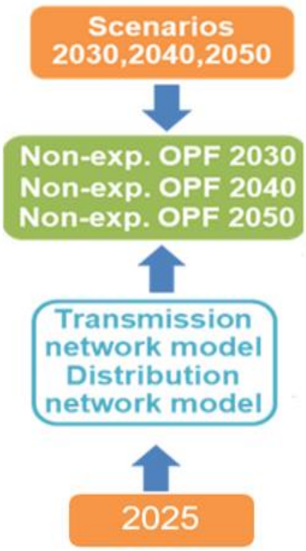

Figure 1-1 Planning process.

The data used for the network modelling is important because it forms the basis for the regional cases and their analysis of the future investment decisions. Both transmission and distribution networks have to be considered in order to provide grid constraints and allow to calculate a realistic generation dispatch of the system, able to highlight grid investment needs. The main modelling difficulty is the availability of data of the physical power systems since this kind of data is usually confidential information and is not available for research use. In the process of building a model, it is necessary to consider the main, i.e. the most essential, characteristics of the system. The six regional cases have been done using realistic geo-referenced models of the corresponding transmission and sub-transmission systems.

The modelling activity includes:

- Review of the data collection for the transmission and the distribution network modelling.

- An overview of the principles used for the adaptation of the Pan-European market study results.
- An overview of the principles used for the localization of load and generation, including those resources which are in distribution networks.
- Networks adaptation to the input data requirements of the new planning tool developed in FlexPlan.

In order to ensure a coherent approach for all cases, it was decided to use a common base dataset. After an initial research, the chosen dataset was the grid model provided by the European Network of Transmission System Operators for Electricity (ENTSO-E) for the preparation of their 10-year network development plan (TYNDP) from 2018. This dataset contains a pan-European transmission system model for 2025, which is the base year for the TYNDP studies. Despite ENTSO-E data support, there were several data gaps (e.g. 110 kV-150 kV sub-transmission network data for several regions, distribution network data, generation and load data, etc), which had to be collected for comprehensive long-term network analysis, considering the transmission and the distribution systems together. No doubt, the initial version of a data set needs to be iteratively refined and supplemented in subsequent project activities.

Particular attention was devoted to the distribution networks. From the planning perspective, distribution systems are expected to be massively affected since the evolution of the generation mix and demand side management is taking place at the lowest voltage levels mainly. However, the collection of distribution networks data (especially considering that the entire territory of each regional case needs to be covered) is critical. For this reason, in the FlexPlan project, real distribution systems have been replaced with simplified synthetic networks that have been created on the basis of statistics extracted from the real systems.

ENTSO-E has developed a methodology for the technical economic appraisal of transmission grid development projects [3], thus ensuring a common framework for the TYNDP projects. However, such methodology leaves many aspects, most notably the environmental ones, as described in a qualitative way. In the last years, there has been an important debate on what aspects could be reasonably translated into a cost parameter so as to be included into an overall techno-economic assessment, what, strictly speaking, is usually denominated Cost-Benefit Analysis (CBA) in order to find out the most beneficial expansion plan while retaining an approach that is both computationally feasible and requesting data that can be easily obtained. Some studies have been led in the hypothesis that some aspects could not easily be reconducted to a sheer cost parameter and this gave rise to the Multicriteria Analysis methodologies. One of the strong points of the FlexPlan methodology is having succeeded in setting up an applicable methodology able to convert all techno-economic appraisal aspects into a CBA (i.e. a “cost oriented” approach as opposite to multi- criteria approaches). Thus, in the FlexPlan methodology, the objective function consists of a maximization of the social welfare of the system. This is obtained by minimizing the sum of network investments, operational costs associated with system dispatch, consumer costs due to curtailment and shifting of load demand, and environmental impact costs. This deliverable provides a description of the costs and economic data sets that will be used for regional cases development within regional case analysis.

Another important issue is the environmental viability of a solution based on the territory the new infrastructures have to cross and the impact of thermal energy production variations in terms of carbon footprint and air pollution. The data of existing thermal power plants that will be still operating in 2030 was used for the environmental analysis for 2030, 2040 and 2050. The air quality impact estimation accounts for existing thermal generators, the carbon footprint analysis generalizes to both investments and operations through the life cycle of each asset. For datasets providing the required technical data (efficiency or fuel consumption and energy production, stack parameters) but lacking all or part of the emission data (the amount of pollutant emitted mass for a unit of fuel), published emission factors have been considered [64]. Missing parameters to be used for plume rise estimation due to thermodynamic buoyancy are filled using values typical of power plants of the same technology and size. A simplified model was used to evaluate air quality impact, based on sensitivity coefficients estimated by a chemistry and transport model (CTM), able to reproduce all chemical and physical processes which air pollutants undergo in the atmosphere. The modelling grids used for the regional cases runs of the CTM model have been chosen to balance coverage and resolution of the output fields on the one hand, and computational efficiency on the other.

1 Introduction

The main objective of this deliverable is to provide a description of the data sets needed for regional case studies over the three FlexPlan target years (2030, 2040, 2050) by the new planning tool developed in FlexPlan.

The data used for the network modelling is important because it forms the basis for the regional cases and their analysis of the future investment decisions. Both transmission and distribution networks have to be considered in order to provide grid constraints and allow to calculate a realistic generation dispatch of the system, able to highlight grid investment needs. As such, this deliverable provides an overview of the data sets required for the network modelling, describes the common modelling approach, the principles for the localization of load and generation, including those resources which are located in distribution networks. The construction of the regional network models includes modelling of transmission and distribution networks and flexibility elements located at both levels of the grid.

The input data, such as the hourly electricity demand, distribution of the renewable energy source (RES) and cross-border flows time series is the result of the pan-European market simulations. Such simulations provided a set of data for each regional case, for each of the three target years, for all scenarios. However, the market results represented the distribution of generation and load at regional level, i.e per province or big city. One of the critical aspects when analysing the network, RES integration and demand fluctuation is to have a sufficient network granularity to explore the contribution of each single flexibility element, which is usually very local. Thus, the regionalized load and generation data provided by the pan-European simulations were further localized by fitting them to more detailed regional grids, also considering distributed energy sources located in distribution networks. This deliverable provides a description of the principles used for localization of load, generation and flexibility elements (energy storage systems and flexible loads), including those located in distribution networks.

After the six regional cases were set up with the relevant scenario data, they were put in the JSON (JavaScript Object Notation) data format according to the input requirements of the new planning tool. This document contains a general description of the main principles used to link the input data required by the new planning tool and the modelled regional networks. Moreover, a detailed description of the structure of the JSON files is provided.

The growing share of RES and, the appearance of new loads such as, electric vehicles (EV) and heat pumps, have caused new grid development challenges in distribution and transmission networks. Traditional network development foresees the construction of new or upgrade of existing grid elements such as transmission lines, transformers etc. FlexPlan aims at creating a new tool for optimizing transmission and distribution grid planning, considering the placement of flexibility elements as an alternative to traditional grid planning. This approach aims at reducing the overall power system costs i.e. infrastructure deployment, operation costs associated with system dispatch, consumer costs due to curtailment and shifting of load demand, and environmental impact costs. The main information regarding the applied planning and optimization criteria is given in deliverable D1.2 of the FlexPlan project [4]. The present deliverable explains the reasons for choosing a “cost oriented” approach versus multi-criteria approach and provides a description of the costs and economic data sets that will be used for regional cases development within regional case analysis.

Additionally, FlexPlan aims also to evaluate the environmental impact, including impact on air quality for thermal generation, carbon footprint, impact on landscape of new transmission and distribution network. This document describes the required environmental data sets and provides a description of how environmental models are set up and used to obtain their environmental parameters for the new planning tool.

The following chapters are organized as follows:

- Chapter 2 presents an explanation of the motivations for choosing a “cost oriented” approach versus multi- criteria approach for the cost-benefit analysis.
- Chapter 3, 4 provide a description of the data sets used for the regional power system modelling.
- Chapter 5 provides a description of environmental parameters and a description of the environmental models.
- Chapter 6 provides details on fuel costs and economic data.
- Chapter 7 describes the approach applied for the adaptation of pan-European market results.
- Chapter 8 describes some ICT aspects and the JSON file structure.

2 Applied planning and optimization criteria

The core drivers for searching the new approaches to transmission system planning are:

- Deregulation of the power system, where generation expansion is not co-optimized with network development as it was the case for old vertically integrated electric companies: thus, from the modelling point of view, generation is a system parameter and no longer an optimization variable.
- Uncertainties: the increasing penetration of RES generation is pushing more and more towards probabilistic models.
- More rapidly changing conditions.

The planning process typically includes scenario development, security analysis and cost-benefit analysis. A first phase for the development of a grid planning study consists in setting hypotheses for generation and load evolution at the target year by creating one or several power system development scenarios. Then, a network analysis has to be carried out, bringing to the identification of the power system criticalities. The alternative solutions could be formulated on the basis of the TSO planning experience that could overcome the detected bottlenecks. In this way, a single line reinforcement might not constitute an alternative whenever not able to solve a given bottleneck, but it could rather be necessary to schedule a series of reinforcements along a given transmission corridor which are able to increase the transfer capacity between two nodes. Once a series of alternatives has been formulated, the final stage is to put in place a cost-benefit analysis to score them from the techno-economic point of view, so as to be able to locate the best investment solution(s). Basically, a cost-benefit analysis consists of an evaluation of the trade-off between the sum of the investments and the achievable reduction of dispatching costs. This aspect is an important part of the cost-benefit analysis, however not the only one. For instance, traditionally the environmental analyses carried out separately, nonetheless are an important part of the investment analysis.

The basic question is what aspects could be reasonably included into an overall technic-economic assessment in order to find out the most beneficial expansion plan while retaining an approach that is both computationally feasible and requesting data that can be easily obtained.

A possible approach to reduce problem complexity consists of converting all indicators into monetary units. In this case, one speaks of Cost-Benefit Analysis (CBA), which is based on transforming indicators into one unit. However, some parameters might not be easily transformed into monetary units (e.g. some environmental or social parameters). In this case, some authors propose the so-called multi-criteria approach (MCA).

Typically, this consists, in synthesis in the following steps [5]:

1) A set of criteria is defined that could be used to classify alternative investment variants. In that, it is important that potentially overlapping criteria are avoided ("double counting").

2) Then, quantitative indicators are provided to quantify the selected criteria. These indicators can be represented either by absolute measurements (indicators) or through a differential measurement with respect to a base case (impact factors).

3) Thereafter, all the criteria indicators need to be converted into one only, possibly a-dimensional, utility value, expressing the level of satisfaction or approval that a single value of the indicator has towards the society as a whole. Typically, a utility value equal to zero expresses no satisfaction, whereas a figure equal to one expresses maximum satisfaction. The function performing this conversion is in general called a utility function.

4) Once all the indicators have been converted into one only utility parameter, all the indicator values relevant to a single alternative may be linearly combined so as to calculate one only ranking parameter attached to this alternative. In general, a weighted linear combination is calculated, making use of a weights vector. This vector incorporates the reciprocal importance (for the public opinion, for the political and/or technical decision-makers, etc.) of one criterion with respect to the others.

From the description above, a few issues emerge associated with the MCA:

- It is not clear how it can be assured that all the main factors are included while avoiding overlapping.
- The quantification through non-homogeneous indicators does not allow to establish a common metric on the basis of which to compare the reciprocal importance of the different criteria.
- The possibility to set-up utility functions is a potentially interesting degree of freedom but in practice, it seems very complicated and, to some extent, questionable, to establish clear and documented criteria to establish the level of public satisfaction in relationship with the indicator's values. Sometimes, this difficulty is overcome by ad hoc questionnaires, but this does not solve the problem: how to check the significance and the completeness of the statistic respondents' sample? How to move the respondents to reflect into their replies their real opinion and not the easiest way out?
- The establishment of a set of weights measuring the reciprocal importance of the different criteria is a subject to infinite debates, that risk to be never conclusive in case interests are at stake.

For all the reasons above, we deem that the CBA is the most feasible approach. For the few aspects for which an economical quantification can be subject to a very high arbitrariness, a mixed approach could also be suitable. However, during formation of the FlexPlan methodology it was realized that an economic quantification can be given to all relevant factors and thus a CBA approach was preferred.

The new FlexPlan planning tool aims at finding out an optimal combination of new grid investments, both in new grid investments and installation of flexibility devices, at the minimum costs. Thus, the objective is the maximization of the social welfare of the system by minimizing the sum of network investments, operational costs associated with system dispatch, consumer costs due to curtailment and shifting of load demand, and environmental impact costs. More information on planning and optimization criteria can be found in deliverable D 1.2 of the FlexPlan project [4].

3 Network data and modelling

3.1 Transmission and sub-transmission network

The six regional cases are simulated using realistic geo-referenced models of the corresponding transmission and sub-transmission systems. In order to ensure a coherent approach for all cases, it was decided to use a common base dataset. After an initial research, the chosen dataset was the grid model provided by ENTSO-E for the preparation of TYNDP from 2018. This dataset contains a pan-European transmission system model for 2025, which is the base year for the TYNDP studies. The dataset and the required process to obtain it was already described in deliverable D4.1 of the FlexPlan project [1]. However, the grid model received from ENTSO-E does not have a complete set of information required for the scope of the project, and three main data gaps were identified:

- Non-existence of sub-transmission systems for major countries, including Spain, France, Germany and Italy
- Non-existence of grid model for northern countries (Sweden, Finland and Norway were absent, while for Denmark the model only included the continental part of the country, in the continental Europe synchronization zone)
- Non-existence of geographic information (and in some cases there was an anonymization of grid node names).

Considering these limitations of the pan-European transmission system model received from ENTSO-E, it was necessary to perform a complementary and complex data collection process which allowed to complete the transmission/sub-transmission grid models for each regional case. This activity included different data sources and aimed at solving the aforementioned three main data gaps.

A summary table of the main network modelling assumptions can be found in Annex 1. The present chapter reports a detailed description of the criteria utilized for data collection, the primary data sources used and the relevant modelling principles.

3.1.1 Iberian Peninsula

The Iberian regional case includes both Portugal and Spain. The two Iberian countries are interconnected and they share a common electricity market in normal operation conditions (when no congestion exist in the interconnections). In addition, Spain is electrically connected to France, through both alternating current (AC) and direct current (DC) lines, and Morocco. The Spanish power system is divided into the mainland system and the island's systems. There is an undersea interconnection between the mainland and the Balearic Islands, but these have not been considered in the network model and are excluded of the analysis performed in the scope of FlexPlan. Due to the low net load demand of the islands, when compared to the continental one, this load was considered as negligible.

The electrical models of the network come from the ENTSO-E model of the EU transmission network, which was obtained through the signature of an NDA. In the case of Spain, it includes 400 kV, 220 kV and

320 kV (HVDC) network nodes. The Spanish sub-transmission network (operated at 110 - 132 kV) was not included in the model, so it was built up from power system maps, publicly available from REE, the Spanish TSO [6] and from OpenStreetMap (OSM) data (transformers between transmission and sub transmission were neither included in the original model). Since the information in the OSM is not complete, the REE map, which is available only in pdf format, is used to generate the 132 kV network.

Once the 132 kV network nodes were defined from [6] and [7], they were included in google maps with their coordinates. Then, the connection among nodes was carried out through line objects, considering approximately the route described in the REE map. These line objects have a length property (calculated from the coordinates), which was used to establish branches' lengths. The characteristic of a typical 132 kV line was used as reference to create all the lines of the model, i.e. all 132 kV lines have the same electrical parameters (average values). The interconnection nodes between the transmission and sub-transmission networks were identified and transformers were included in the model at those locations. Again, typical transformers for 400/132 kV and 220/132 kV voltage levels were considered. To analyse the network, the model was coded in PSS/E (.raw) file format, and power flow has been run considering the generation and loads included in the model received from ENTSO-E (single snapshot). The power flow calculation converged.

For transmission model nodes, the location (coordinates) obtained from PyPSA [9], which was originally identified as an alternative data source for grid location information, were not accurate and, in some cases, we observed a big mismatch (see Figure 3-1- Figure 3-3).

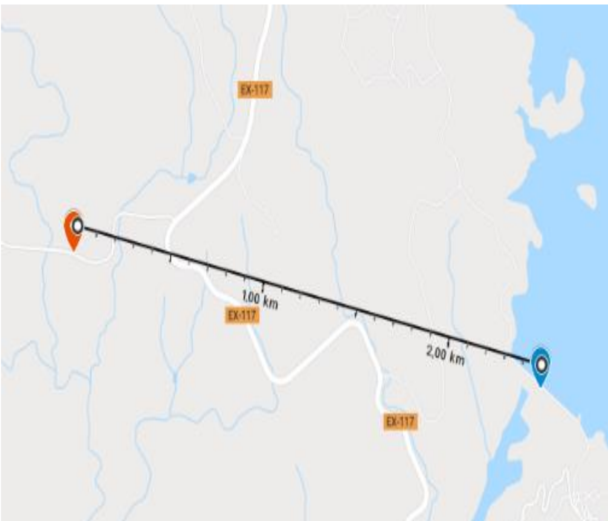

Figure 3-1 - Location mismatches from PyPSA (GridKit) JM Oriol, 2km mismatch.

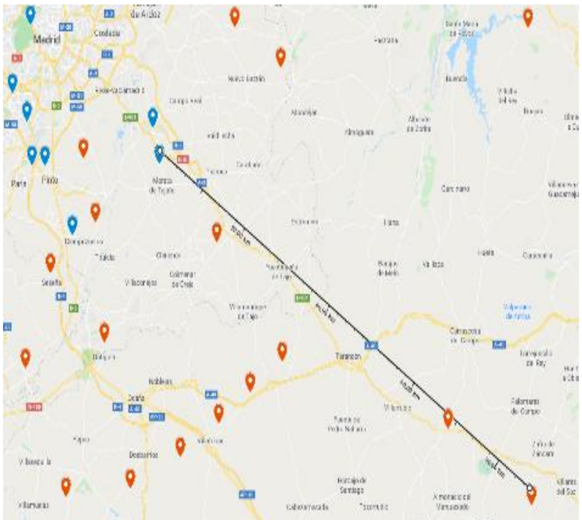

Figure 3-2 - Location mismatches from PyPSA (GridKit) Morata, 80 km mismatch.

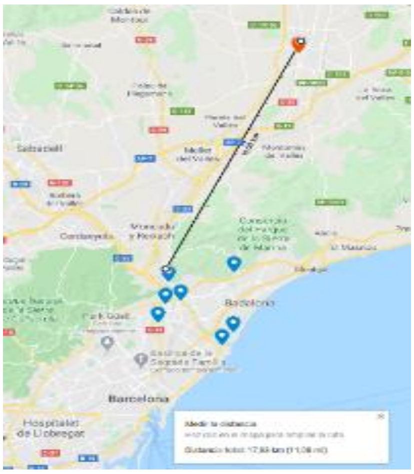

Figure 3-3 - Location mismatches from PyPSA (GridKit): Gramanet, 18km mismatch.

To overcome this, we used the satellite image from Google Maps to identify substations. From 1229 nodes in the Spanish transmission model (sub-transmission not included), we found the exact location for 88% of the nodes and, for the remaining ones, we included an approximate location.

In the case of sub-transmission, the locations from OSM [7] were quite accurate, so we had just to add the nodes that were not included in the OMS but we considered from the REE network map. This was done, again, with the help of Google Maps.

For Portugal, the model shared from ENTSO-E contains all relevant voltage levels, namely: 400 kV, 220 kV and 150 kV and additionally all transformers for the 63 kV (TSO/DSO interface) are also considered. The Portuguese dataset was then complete, and the main action performed was related to the need of identification of grid node names as the dataset contained anonymized node names (coded with numbers). The model received from ENTSO-E was directly convergent. Portuguese transmission network, after a first process of identifying grid nodes (as they were anonymized in ENTSO-E model), the identification of corresponding locations was obtained with the collaboration of REN. All provided node locations are exact, with the exception of some switching stations, which are not relevant for the study considered in the scope of FlexPlan. The Portuguese TSO, REN, contributed in the identification of the location of existing grid nodes and contributed as well in the process of de-anonymizing the received dataset from ENTSO-E.

A characterization of the nodes was performed indicating if it was a substation, generation or industrial load connected directly to the transmission system. In the case of generators and loads, their type was identified. In the case of generators also the size and year of installation.

The Spanish sub-transmission (132 kV) network is not complete, in the sense that not all the existing substations, generation or loads were included in the model. A selection was carried out trying to keep the grid structure, but using a lower number of nodes. The lines and transformers included in the network model at this voltage level come from single line generic line type and two generic transformer models (400/132 kV, 220/132 kV). These generic models were obtained as an average of real lines and transformers at that voltage level. The generic line was defined in per unit values and the final electrical parameters of each line were calculated using the length extracted from the Google Map measured between the connected nodes. The modelled network is a simplification of the real one, since not all the nodes and lines are included. Nonetheless, it provides a realistic representation of the existing network, as most relevant lines were selected and a preliminary power flow analysis does not result in large deviations from known values using full models.

In Spain, the names of the nodes follow the names included in REE maps, plus information of the voltage level to assure that they are unique. Special characters from the Spanish language, such as the "ñ" and accents, were eliminated.

The transmission and sub-transmission network nodes are represented in a Google Map, through their location. In the case of sub-transmission, the lines are also depicted in the map Figure 3-4.

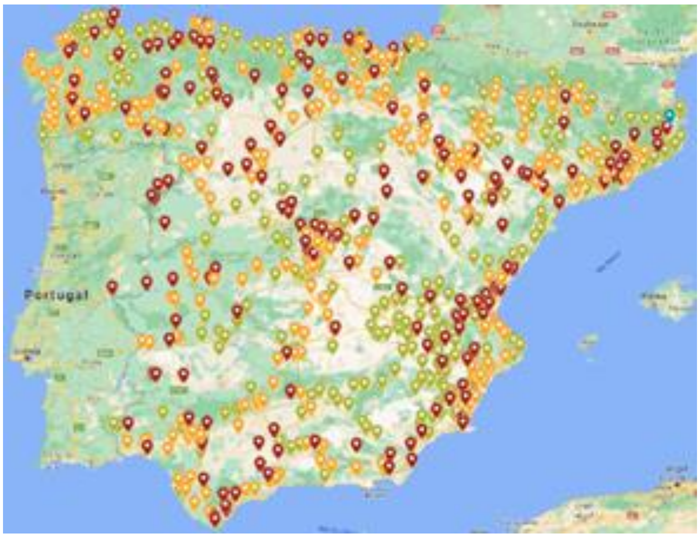

Figure 3-4 - Location of 400 kV, 220 kV, 132 kV and 320 kV DC network nodes in the map of Spain. Right: zoom showing the 132 kV lines.

In the case of Portugal, a similar approach was followed, eliminating possible special characters and simplifying the name of some substations which nonetheless keep a name very similar to the official one.

3.1.2 France and Benelux

The main data for the transmission network modelling was downloaded from the ENTSO-E TYNDP data. The following voltage levels (in kV) can be found for Benelux and France:

- Belgium: 380, 220, 150, 70, 36, 33, 20, 18, 16, 15, 12, 11, 6.
- Netherlands: 380, 220, 150, 110, 66, 52.5, 52, 50, 34, 33, 27, 25, 21, 20, 19, 16, 15, 11, 10.
- France: 380, 225, 150, 45, 42, 20.
- Luxembourg: 220, 15.7, 13.7.

While Belgium, Netherlands and Luxembourg had a complete grid model, the French grid was incomplete. The French data contains transmission substations (V >=150 kV) and some low voltage networks related to the low voltage side of generation units (V =< 45 kV), while data were not available for sub-transmission network (63 kV and 90 kV) which connects distribution to the transmission network. Openly available data [10][11][12] was utilized to derive the sub-transmission network and merge it to the French grid.

The modelling data for Benelux and French grid was provided initially in the PSS/E (.raw) format. The data was converted to PowerModels.jl [13] for testing using an Optimal Power Flow (OPF) and further processing.

As aforementioned, the data set for the French sub-transmission network was incomplete. For providing the sub-transmission network of the French grid, the ENTSO-E data and open-source data were compared in order to import the sub-transmission data from the openly available data to the ENTSO-E data. Not only that, but also new transmission-level substations were added to ENTSO-E data. This process is explained briefly below:

For the electrical network modelling of the added AC lines, cables and transformers average electrical parameters were used which were taken from the TYNDP data set as well as from the literature. For sub-transmission network modelling, ENTSO-E data and open source data were compared in order to supplement ENTSO-E data with the data from the open sources. Thus, new transmission substations were added to ENTSO-E data. This process is explained briefly below:

- Transmission network:
  - Name inconsistency between the substations of the two data sets, were solved by a fuzzy lookup method in Microsoft Excel and the rest manually.
  - Voltage inconsistency between two data sets was resolved. While the open source data has 380 kV, 225 kV and 150 kV, ENTSO-E data has 400 kV, 380 kV, 320 kV, 225 kV, 220 kV and 150 kV.
  - The open-source data contains three sets showing the overhead lines (OHL) and underground cables (UGC) on the 63 kV and 90 kV levels, respectively as well as a data set with substation names. For the OHL and UGC connections, the connecting substation names are provided in the respective data sets. However, some substations that exist in the OHL and UGC data were missing within the substation data set. As such, an algorithm has been developed to import the missing substations from the OHL/UGC data sets to the substation data set.
  - There were some transmission-level substations in each data set that did not exist in the other source, e.g., open-source data vs TYNDP data set. It means that none of the transmission-level substation data of two data sets is complete alone, but they are complementary. New transmission-level substations were added thus to the TYNDP data set.
  - The transmission substations imported from the open source data cannot be isolated from the substations in the TYNDP data and thus need to be connected properly. There can be three possibilities: 1) The substations from the open source data are connected via transformers to the TYNDP substations and both are at the same location but on different voltage levels 2) They can refer to “dummy” substations, e.g. line splitting within HV lines of the TYNDP data set 3) they are actual substations not present within the TYNDP data set and are connected via actual

HV lines. Different algorithms were developed to distinguish between the three possibilities and corresponding elements were created to connect all substations coming from the two different data sources. As such, many new transmission branches have been added to the TYNDP data.

- Sub-transmission:
  - The missing 90 kV and 63 kV substations were added to TYNDP data.
  - An algorithm was developed to create the correct sub-transmission branches in forms of transformers and lines.
  - For the electrical network modelling of the added AC lines, cables and transformers average electrical parameters were used which were taken from the TYNDP data set as well as from the literature.

The ENTSO-E TYNDP data set does not provide any geographical data including latitude and longitude of the substations. Some of the substation locations were imported from open-source data, but the remaining 500 substations were found manually using maps.google.com and openinframap.org.

The resulting grid model of the French transmission and sub-transmission grid can be seen in Figure 3-5 below. Reasonable approximations were used for the geographical locations of some substations.

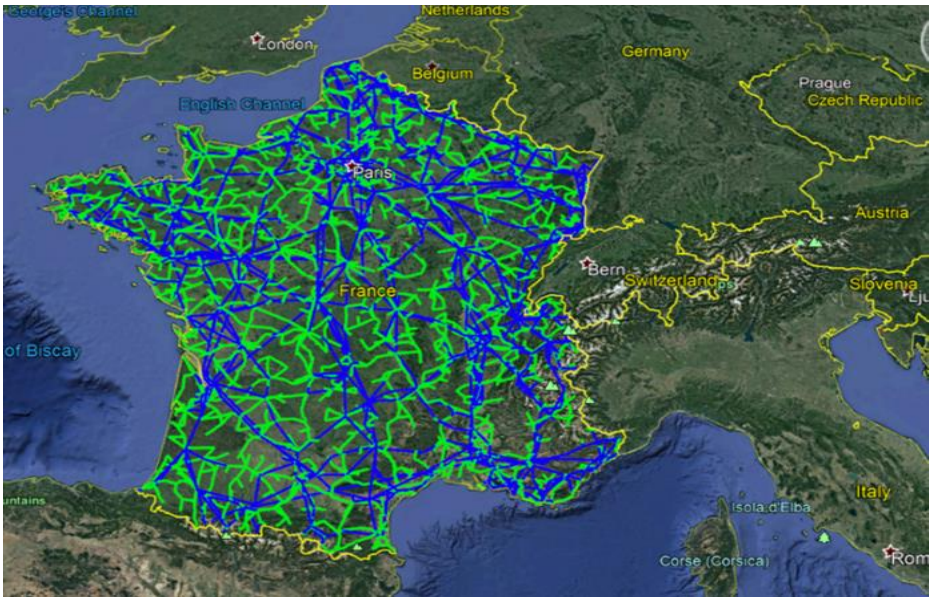

Figure 3-5- French transmission and sub-transmission networks

Since not all the substation's geographical locations can be found on OSM, some of them could be spotted based on their location on Google Maps and some of the substations were approximated based on their adjacent substations with known geographical information.

3.1.3 Germany, Switzerland and Austria

The main data for the transmission grid modelling is depicted from the 2025 grid model provided by ENTSO-E. The sub-transmission grid, consisting of the high-voltage-level (110 kV level in Germany and to some extend 66 kV in Switzerland/ Austria), is based on OSM Data.

The modelling is carried out using MILES (Model of International Energy Systems)[80] and PowerFactory. For the electrical grid modelling (AC, DC lines, transformers) the electrical parameters of the 2025 grid model provided by ENTSO-E are used. If parameters were not available or implausible in comparison to existing MILES grid data, an analysis based on open sources [14] the German Grid Development Plan [15] and TYNDP [16] is carried out for different years to determine the modelling parameters. If coordinates are missing, the internal MILES database and the ENTSO-E map are used to add the information.

The sub-transmission grid (110 kV) is generated with the help of OSM information [17]. The algorithm developed uses stored information on lines ("power = line" and "power = cable" with "voltage = 110000") and on substations at the high-voltage level. Due to the large number of high-voltage grids in the Regional Case, an automated correction of the stored information in OSM is necessary. This includes, for example, consistency checks for line corridors, so that the number of parallel lines does not change in the course of a route. The nodes of substations are set to the centre of the specific area, for which the polygons of the boundary sometimes have to be adjusted. Tie-off points have been also made plausible, to get the correct high voltage grid topology (as OSM data does not provide information for the cables at the substations). The connection points between the 380/220 kV grid and the 110 kV grids are depicted using the location of 380/220 kV stations as well as information provided by OSM-users [18].

The geographical location is derived by using the ENTSO-E Map and an internal MILES-database.

The switching state of the 110 kV grid groups can only be considered in approximate terms. If some 110 kV-groups are permanently connected during operation, this is not displayed. High-voltage grids (means the voltage higher than 110 kilovolts, AC, 50 HZ, but which does not exceed 380 kilovolts) are usually meshed and have interconnection points to the extra-high voltage (EHV, means the voltage, which exceeds 380 kilovolts, AC, 50 Hz) level. High-voltage grids are operated in groups that are not directly connected at the EHV-level. This grouping is ensured via switching positions at the respective substations (often interconnection points to EHV). However, this switching state is not part of the OSM data and therefore has to be assumed by analysing the overall grid topology.

3.1.4 Italy

The main data for the transmission network modelling was downloaded from the Ministero della Transizione Ecologica (Ministry of Ecological Transition) website [19]. The obtained geographical network model of transmission, 200 kV- 400 kV (including sub- transmission 110 kV -150 kV), dating back to 2008, was updated to the 2020 reference year by using available information from [7][20]. The obtained data sets include the list of the grid nodes names with coordinates, voltage (kV), type of substations, length of the lines (km).

The modelling work was carried out with MATPOWER, with data modelling scope and requirements taken from [21]. For the electrical network modelling (AC, DC lines, transformers), average electrical parameters were used, drawn from the 2025 network model provided by ENTSO-E. Some standard parameters were taken from public sources. In a first set of simulations, power capacities of all lines (AC, DC) and of transformers were not limited, in order to allow to perform a maximum power flow calculation. Subsequently, the capacities were adjusted according to the maximum values from the yearly calculation results, taking possible physical limits into account. For the modelling of cross-border interconnections, ENTSO-E data were used, as well information available on the Italian TSO website. The interconnection capacities were checked and agreed with representatives of the neighbouring regions. Load and generation were modelled at nodal level (without transformers). Generation was modelled according to installed capacities, voltage level and geographical location of the power plants. The information about national load distribution was gained from the previous RSE studies (a description will be given in chapter 4).

The reference case model was tested and analysed by MATPOWER. Actual 2020 time-series for load, generation and cross border exchanges were downloaded from ENTSO-E transparency platform. Thanks to the availability of data with bidding zone granularity, the regionalization of total load and generation was performed per zone. (The generation, consumption of each node in all zones was calculated as a percentage of the total generation and consumption in its zone). Finally, by using the actual time-series and applied calculated share of the total generation, the consumption per node were estimated for all hours of the considered year by using MATPOWER. In order to assign power capacities to lines and transformers, the OPF simulation results were used. The lines and transformers loading statistics were compared with available physical parameters and then, the power capacities were assigned accordingly taking its limits into account.

Since the primary source of information consists of maps [19][20], the obtained network nodes were easily associated with their geographical location (coordinates).

3.1.5 Balkan region

The main data for the transmission network modelling was downloaded from the ENTSO-E TYNDP Data Set. The obtained network model of the transmission 110 kV - 400 kV network was for 2025, it was updated to 2030 reference year by using available information from [7][20] and an internal EKC database. The obtained data sets [7], [20] included the list of the nodes' names with coordinates, voltage (kV), type of substations, length of the lines (km).

The transmission network model of each country from the regional case was downloaded from ENTSO-E Data Set as CGMES (Common Grid Model Exchange Standard) and converted to PSS/E (.raw) file format. Since the models were for 2025, they were updated to the 2030 reference year in terms of generation capacities. Power flow simulations were carried out for each model with PSS/E. Some of the models initially diverged (Serbian). The models of all countries were merged into a common grid model of Balkan region. The interconnection capacities were checked and agreed with representatives of the neighbouring regions.

Since the primary source of information consists of maps [7], [20], some of the network nodes had their geographical location (coordinates). For the rest of the nodes, geographical location (coordinates) have been taken from EKC's internal database.

The model downloaded from the ENTSO-E Data Set contained a 110 - 400 kV transmission network of the entire region with all the necessary parameters so that no simplifications were necessary.

3.1.6 Nordic region

Since the ENTSO-E data set used for the transmission grid for other regional cases did not include data for the Nordic synchronous system, grid data had to be sourced elsewhere. The four countries have different policies with respect to sharing of power system data. NVE, the Norwegian Water Resources and Energy Directorate [22], has shared with SINTEF Energy Research the data of the Norwegian transmission grid under a Non-Disclosure Agreement. Therefore, the Norwegian transmission and sub-transmission grid is fully covered with highly reliable data. In the case of Denmark, the Danish TSO Energinet releases the transmission system data through an Excel file openly on their website [23]. Finally, data from Finland and Sweden is not available through their institutional websites, neither there has been the opportunity to have this data available through direct contact to the respective TSOs. For this reason, in order to have a complete model of the transmission grid of the Nordic region, the availability of open source data has been fundamental. The main sources that have been used are PyPSA-EUR model [24], and OpenStreetMap (OSM) [7].

The Norwegian grid model includes 2148 buses, primarily at voltage levels 132 kV to 420 kV, but also around 150 buses at lower voltage levels (10 kV to 66 kV) that each represent an individual distribution grid.

The Danish grid model includes 300 buses, primarily at voltage levels 132 kV to 400 kV. The data set includes partial information about generators and loads connected to these network levels. The data set comprises a balanced power flow solution for the year 2020.

The grid model for Sweden and Finland is a grid model combined from PyPSA-EUR data for voltage levels 220 kV and above and OSM data for voltage levels from around 100 kV to 130 kV. (The original dataset includes also buses at 15kV, 20 kV, 45 kV and 70 kV, which are excluded from the model.) The full transmission and sub-transmission grid model for Sweden and Finland consists of around 3200 buses.

OSM data were used for modelling the sub-transmission grid of Sweden and Finland by exploiting the Overpass API using the OSMnx Python package [25][26]. A topological investigation of the grid models generated returns a single island for both countries. OSM-based data was first converted into a graph, and finally into a pandapower network, where the line sizing was based on pandapower standard lines libraries. Unconnected buses were excluded by the power network. The resulting network was analysed in terms of connectivity by exploiting pandapower topological functions, in order to guarantee that the network is defined by a single island fully connected.

Partial verification and validation of the grid model and the methodology for creating it was carried out through visual inspection and comparison with maps through OSM web interfaces and maps published by Nordic TSOs.

The grid models for Sweden and Finland are generated based on maps and therefore include information about the geographical location of all buses. The data set for Norway does not include information about the geographical location of buses or the length of lines. Neither the data set for Denmark includes information about the geographical location of buses, but coordinates were obtained with an automatic script that matches Energinet bus names with the corresponding OSM stations names, integrated with manual inspection. The methodology is based on work previously done by SINTEF in the SmartNet project [28].

The provided data set from the Norwegian TSO does not contain geographical information about the busses. To remedy this, a combination of several sources where geographical information is included, has been used to map each of the busses connected to the grid. These sources include a database of transformers, a database of power plants (and SINTEF in-house data about which busses connect to which power plants) and geographical data for an earlier iteration of the grid model (but a lot of the bus names had been changed). The remaining busses were mapped manually based on their name, and the location of their neighbouring busses. The source for the transformer and generator geographical data was NVE's publicly available grid map [27].

Standard electrical parameters were used for branches (transformers and power lines) in Sweden and Finland; this is the case for the branches from PyPSA-EUR, and similarly standard parameters were applied for the branches generated from OSM data. Both PyPSA-EUR and OSM data sources were converted into pandapower, therefore the line sizing has been based on standard libraries from pandapower. The line size association has been based on the line terminals voltage levels. Since the line's capacity information is not given by the data sources, the line with the highest power capacity among the lines with corresponding voltage level is selected and assigned.

3.2 Distribution network

From the planning perspective, distribution systems are expected to be massively affected since the evolution of the generation mix and demand side management is taking place at the lowest voltage levels mainly (medium/low voltage). However, as anticipated in deliverable D4.1 of the FlexPlan project [1], the real data collection of distribution networks (especially considering that the entire territory of each regional case needs to be covered) is critical. For this reason, for the FlexPlan project, a realistic synthetic distribution systems have been created for each regional case and connected to the transmission network described in the previous section. The adopted synthesis procedure is described in [47], while the source of information (from which the country dependent network characteristics have been extracted) are described within the next subsections.

Thanks to time series returned by MILES simulations, the aggregated power profiles of distributed load and generation have been extracted for each primary substation, assumed to be point of common coupling between high voltage system and medium voltage distribution network. Consequently, these power profiles have been randomly split and spread over the synthetically generated distribution networks to model (for each node) the connected loads and generation units.

3.2.1 Iberian Peninsula

As to Spain, no public data is available for the distribution network. In addition, there are five major DSOs and many small ones and they have historically evolved in different ways. This makes distribution networks to be quite heterogeneous, including the use of different voltage levels. Currently, grid development is carried out considering common procedures at national level, but this was not the case some years ago. Availability of a number of distribution network data sets from previous projects permitted to extract some characteristics of them. These network data are not public, so they were not shared and only non-confidential statistical parameters have been calculated according to the approach described in [1]. These statistical values are the input for the creation of synthetic distribution networks. Some adaptation of the input data had to be done to permit the script to extract the characteristics parameters of the network. Only connection nodes between transmission and distribution networks have been considered and their locations were identified.

The FlexPlan consortium includes e-distribución (one of the major Spanish DSOs) as a linked third-party. e-distribución provided statistics on distribution network characteristics, which agrees with the ones already collected by RSE from the information reported above.

For Portugal, as the full map of the high and medium voltage networks for the distribution system is available in [29]. Using this map and the information annexed to the same document, the actual topology and characteristics of the network has been reconstructed and will be used in the context of FlexPlan. All primary substations of the distribution network also have their location identified. A network model has been built, using the real grid topology. The grid model was created using public data shared by E-REDES, the major Portuguese DSO.

3.2.2 France and Benelux

Multiple networks have been downloaded from DiNeMo [30][31]. Depending on the location of the primary substation (transmission or sub-transmission substation), a distribution network is assigned to the substation distinguishing the rural, suburban and urban nature of the substation location. Eventually the individual demands of the distribution network are adjusted such that the total demand of the distribution network corresponds to the nodal demand which has been obtained by disaggregating the nodal demand provided by MILES.

For the distribution networks, the most used voltage levels in Belgium and the Netherlands are 30 kV, whereas in France 20 kV and in Luxemburg 15 kV is utilized most. Thus, the created networks used these voltage levels.

Starting from this information, the statistical values requested by the procedure described in [1] have been extracted and provided as input for the creation of synthetic distribution networks which cover the entire regional case territory.

3.2.3 Germany, Switzerland and Austria

The distribution grids of this Regional Case are not publicly available. However, data from SimBench [78] can be obtained to use the synthetic network generator. In SimBench, synthetic and representative network models were derived based on a detailed literature review and using developed tools [32]. The supply situation in Germany is analysed for the medium and low-voltage grids. Clusters are formed on the basis of municipal data, such as population density and community area [33]. For these clusters, which also allow a differentiation into urban, semi-urban and rural regions, grids are subsequently generated. In the case of the low-voltage networks, the networks are generated with the help of OSMoGrid, an OSM tool [34] and subsequently analysed, by comparing the generated grids to existing distribution grid data and working with DSOs that were part of the SimBench-project. The medium-voltage grids are generated using average properties of the grid areas. In summary, SimBench provides network data and evaluations for different types of municipality [35].

This information is then used to allocate low and medium voltage grids to the various transmission grid nodes. The municipality information [36] can be used to assign the appropriate medium and low-voltage grids. The methodology for integrating the generators and loads regionalised by MILES [80] in the distribution grid is presented in the respective following chapters.

The collected data is based on the results of the SimBench research project in which TU Dortmund University participated. DSOs did not participate in the network modelling and validation activities in FlexPlan, but DSOs were associated partners of the SimBench project. Starting from the available models, the statistical values requested by the procedure described in [1] have been extracted and provided as input for the creation of synthetic distribution networks which cover the entire regional case territory.

The geo-coordinates of the transmission grid and the substations in the transmission grid can be used to derive the starting points for the subordinate medium and low-voltage grids.

3.2.4 Italy

No public models of the Italian distribution systems are available and synthetic networks were created using several sources of publicly available information [37]-[44]. The analysis of the collected data allowed the adoption of a synthetic network generator [45], capable of providing a number of different distribution grids covering the entire territory of the regional case.

Voltage levels and connection with the transmission network has been assumed on the basis of the Italian region/province, power transfer capacity of high voltage network, density of final users. In particular, the number of primary substations (point of common coupling between transmission and distribution) per province is publicly available information [46] and the nominal power of their transformers has been hypothesized from the analysis of:

- Development plans released by Italian distribution system operators [37].
- The electricity consumption in each Italian province published by Terna [38].
- Geographical concentration level of final users released by Italian Regulator [39].

Thanks to the data reported in [39], also the characteristics (urban, rural, etc.) of distribution networks have been gathered. This analysis returned the amount of distribution grids connected to the transmission system for each province, together with their rated power and geographical extension. Adjustment (fine tuning) of the obtained distribution networks has been performed [47] by comparing their simulation results with the outcomes of previous studies reported by national regular and network operators.

The process described in the previous section identified the amount and characteristics of distribution networks for each Italian province. Thanks to this information and from the analysis of the Italian transmission network model described above, the precise geographical location of each primary substation has been defined. In fact, the transmission network model extracted from [19] included information concerning the geographical location of distribution transformers.

Concerning the geographical location of distribution nodes, no sources of information have been found. For this reason, no geographical coordinates are assigned to electrical nodes belonging to radial distribution grids (except for the primary substation).

During the previous collaboration experience between RSE and Italian DSOs, models of real distribution networks have been collected. These networks have been analysed and the extracted high-level statistics have been used in order to feed a synthetic network generator [47]. In addition to real networks, information concerning network characteristics have been extracted from:

- Textbooks, literature and presentations on the Italian distribution systems [40][41][42].
- Distribution grid code released by e-distribuzione [43].

- Documentation released by the Italian Regulator[44].
- Rules defined by the Electrotechnical Standardization Committee [44].

The FlexPlan consortium includes e-distribuzione (one of the major Italian DSOs) as a third-party partner. e-distribuzione provided statistics on distribution network characteristics, which agrees with the ones already collected by RSE from the documentation reported above.

3.2.5 Balkan region

No public models of distribution systems for Balkan region are available and synthetic networks were created using modest sources of publicly available information [30][44]. The analysis of the collected data allowed the adoption of a synthetic network generator [45][47], capable of providing a number of different distribution grids covering the entire territory of the regional case.

Voltage levels and connection with the transmission network has been assumed on the basis of the power transfer capacity of high voltage networks. In particular, there are only 4 nodes in Balkan region at which industrial consumption is connected. For all other nodes, we assumed that there is a distribution network connected.

Based on EKC internal database and knowledge of the region, separation of the rural and urban zones has been made.

The DiNeMo tool [26] has been used to obtain parameters about the distribution networks in the region. These data have been used to feed a synthetic network generator [47]. Information concerning network characteristics required as input data for DiNeMo tool has been taken from EKC's database and some public sources [48]. Starting from this information, the statistical values requested by the procedure described in [1] have been extracted and provided as input for the creation of synthetic distribution networks which cover the entire regional case territory.

From the analysis of the transmission network model for Balkan region described above, the precise geographical location of each primary substation has been defined. In fact, the starting transmission network model also included information concerning the geographical location of distribution transformers (distribution transformers are referring to the transformers in primary substations 110/35 kV, 110/20 kV, it was considered all transformers in primary substations between distribution and transmission network as distribution rather than transmission transformers). Concerning the geographical location of distribution nodes, no sources of information have been found. For this reason, no geographical coordinates are assigned to electrical nodes belonging to radial distribution grids (except for the primary substation).

3.2.6 Nordic region

Due to lack of detailed distribution network data, a synthetic distribution network generation method developed by RSE is used for the Nordic region [45] (the detailed procedure is described in [77]). Parameters extracted from real networks from Nordic Region are to be used for generating synthetic distribution networks. (As of 2021-06-16, parameters are extracted for two Norwegian 22 kV urban networks obtained by SINTEF and a data set for Denmark with 10.5 kV rural networks that was previously obtained by RSE.)

Municipality level population projections for 2030 is used from the following sources:

- For Norway: Statistisk sentralbryå [49].
- For Finland: Tilastokeskus [50].
- For Sweden: Statistics Sweden [51].
- For Denmark: Statistics Denmark [52].

For Norway and Denmark, information about load and generation in the grid models is used to determine nodes in the transmission and sub-transmission system to which (synthetic) distribution grid models will be attached: Transmission and sub-transmission nodes to which only generation is assigned (or neither load nor generation is assigned) in the grid models will not be assumed to connect to a distribution system. For Norway, bus names and information about the localization of major industrial loads will also be used to select some load buses to which no distribution grid is assumed to be connected.

A method has been developed internally in SINTEF to designate transmission buses as either rural or urban. First, all transmission buses are mapped to municipality borders available on OSM and with conditions of the population density the designation is carried out. A threshold of 150 inhabitants per square kilometre is used to designate a rural and urban distribution network.

4 Generation and load input data and its localization

4.1 Generation

The simulation of the six regional cases designed in FlexPlan requires the existence of multiple data sets, which allow to simulate different energy scenarios, using the grid datasets described previously. An important activity to complement the existing grid models was dedicated to the collection of data related to generation units. This data is required to characterize dispatchable and non-dispatchable generation and also to allow allocating Pan-European market results (installed capacities and generation time series) to specific power generators in all analyzed horizon years: 2030, 2040 and 2050.

The collection of data for generation units was already started in the context of pan-European market analysis, as this data was required as an input for the MILES tool. In Deliverable D4.1 of the FlexPlan project [1] a short description of data sources is performed, regarding the creation of a database for generators. However, given the scope of Pan-European market analysis, the work was mainly focused on the identification of existing generation units, as this data was used in the regionalization process followed by MILES. In the currently reported work, the focus is on the identification of all necessary power plants required to match the necessary total installed capacities and time-series generation profiles, starting from the outputs of market analysis (regional level) up to the nodal level required in the simulation of the regional cases and representing the created pan-European energy scenarios.

The different regional cases had to follow different procedures in order to complete the identification and allocation of installed capacities. These are linked to the quality and availability of previous existing data (as the grid model received from ENTSO-E for a 2025 scenario). In the next sections, the main methodological rules and principles are given for each regional case.

The installed capacities are pre-given by ENTSO-E scenarios for each target year, the description of the principle for an allocation of the power plants for 2030, 2040, 2050 in the grid model is given below in the following sections for each region. The MILES provides generation data, containing the distributed generation time-series on higher voltage levels of transmission substations. The connection to the actual distribution node is based on a pseudo random allocation algorithm, managed by the synthetic network generator [47].

For the allocation of the new, perspective power plants for 2040 and 2050 the same approach as for 2030 will be used, mainly by using the current location of the existing connections and considering the location of natural resources and local restrictions.

4.1.1 Iberian Peninsula

For Spain, the first identification of power plants came through the analysis of network nodes in the ENTSO-E model. The identification of the nodes, including the satellite view and the information included in the REE power system maps, permitted classification of the nodes per type, including the generation type.

For dispatchable generation, which are thermal and hydro power plants, technology, installed generation capacity and construction year were found from available information (several sources including utilities reports, [84] and [85]). According to the National Integrated Energy and Climate Plan (PNIEC), it is expected that the installed capacity of thermal power plants will not change much in Spain. For hydro power plants, as well, reservoir volume was considered to model them as storage [86]. Inflows per basin were considered and disaggregated to power plant level [87]. Biomass plants were found on the internet one by one (several sources). Wind and solar plants were only identified when their substation was connected directly to transmission or sub-transmission networks. To make the scenarios coherent, the input data from MILES [80] was checked and compared with data from the created grid model, taking the real 2020 installed generation capacity into account. The results are shown in the following Table 4-1(these differences were considered as acceptable):

MW	MILES	FlexPlan model	Difference
Photovoltaics (PV), GW	37-135	37-135	0
Wind, GW	45-86	45-86	0
Hydro RoR	3640	5233	1593
Hydro Res	10975	8513	-2462
Other RES	2226	2806	580
Nuclear	2716	3245	529
Gas	25761	24574	-1187
Pumped storage	9520	9625	105
Storage	9520	9625	105
Total hydro	24135	23370	-764
Total fixed Generation	54838	53996	-842

Table 4-1- Comparison between TYNDP scenarios from MILES and the scenario defined for the project for Spain.

In the case of Portugal, the list of power plants existing in the ENTSO-E data model was considered, for the identification of most dispatchable power plants (thermal and hydro). The list of all existing dispatchable power plants was completed with the help of REN. In the case of renewable power plants (wind, PV, biomass, hydro), the existing power plants were identified by comparing the data from the ENTSO-E model and the data available in [53].

For Spain and Portugal, generation technologies suitable to be connected to distribution networks are Wind, PV, hydro and Other-RES (biomass and biogas mainly). From MILES, generation was allocated to transmission network nodes. The first step was to disaggregate that generation to sub-transmission nodes. For that purpose, each sub-transmission node was assigned to a transmission node based on proximity.

Then the generation power and energy allocated by MILES to one transmission node was distributed uniformly between all sub-transmission nodes assigned to it.

For wind and solar the disaggregation was performed as follows.

- If the transmission or sub-transmission node has no substation type, the wind or solar power plant was connected to the lower transmission/sub-transmission node. If the node has substation type:
    - Solar: if the generation capacity was lower than 50 MW or if it was higher, but is within a big city, it was connected to a distribution node. If not, it was connected to the lowest voltage to the transmission/sub-transmission node.
    - Wind: if the generation capacity was lower than 15 MW, it was connected to the distribution node.

In the case of hydro, the scenario data provided by MILES is assigned considering the following:

- If there was already a dispatchable hydro power energy storage system at a certain node, it was checked if the installed capacity power provided by MILES for that node was higher than that of the defined energy storage system: in this case, the excess power was considered as a new non-dispatchable plant (generator).
- If the generating capacity of a plant provided by MILES was lower than the existing dispatchable capacity of the plant, the existing capacity was considered for that node.
- For the nodes where there were no non-dispatchable or hydro plants, the information from MILES was considered. If MILES indicated that there is hydro and that node had not substation type, the plant has been connected there, if the node had the substation type, the power plant was re-connected to the distribution node. Capacity of 50 MW was not considered a limit, and it was considered that the installed power provided by MILES could be the aggregation of different plants amounting that total size.

For Other-RES, mainly biogas and biomass:

- Biomass plants have been identified in the grid model with their installed capacity. For other technologies falling into this category, such as biogas or ocean energy, it was suggested that new plants are or will be of smaller size and that they will be quite scattered around the territory, therefore, it was assumed to use MILES data. MILES information was checked node by node and if there was a biomass plant, it was considered, if there was not any biomass plant, a new generator was modelled. Some known big generators were connected to the transmission and sub-transmission network and the rest were connected to the distribution network.

When more than one voltage level was available for a certain location, the generators were connected to the lowest voltage level (among 400 kV, 220 kV and 132 kV).

The scenario depicted for 2030 is quite close to current one for conventional plants including hydro. Therefore, new plants are mainly associated with wind and solar (see Table. 1.2). In this case, to be consistent with the MILES data, this information was considered as input for all the existing plants in the future scenarios and it was distributed throughout the network model in accordance with the description in the previous section.

4.1.2 France and Benelux

The different generators are distributed using different influencing factors:

PV generation: in northern countries with low solar irradiation, it is assumed that most of the PV generation is rooftop PV and hence is distributed in urban areas. In countries with high solar irradiation, it is assumed that there are more ground mounted plants, hence they are distributed in non-irrigated arable land.

Wind generation: considering the distribution of existing plants, agricultural areas, and reciprocal to population density.

Hydro generation: upscaling of existing plants and the distribution is divided to the nearest nodes:

For the preparation of the final data sets the following approach is used. For PV generation it can be assumed that it will be mostly rooftop PV. As such, the PV generation will be distributed among the sub-transmission and distribution nodes of the system, making use of the geographical location of the high voltage transmission and the sub-transmission networks. For each high voltage substation, an area of influence will be defined, and the PV generation will be distributed linearly among all sub-transmission and distribution networks located in that area of influence. As most onshore wind power plants are connected to the sub-transmission network, the same approach is going to be applied for them. As the amount of hydro generation is limited, no disaggregation will be performed. Conventional power plants currently under construction will be added to the data sets manually.

4.1.3 Germany, Switzerland and Austria

The main sources for generation data are the German Bundesnetzagentur (The Bundesnetzagentur is a German federal authority that maintains, promotes, and monitors competition in regulated network markets. In its role, it also provides the public with various data on the electrical energy system). Large generation units are maintained in a power plant list [54]. This list enables direct assignment to the connection nodes. Renewable Energy Sources and smaller generation units are part of the so called Marktstammdatenregister [55]. Since an exact allocation to distribution grid nodes is not directly possible with this data and is not practical anyway, only the connection levels are evaluated in order to then locate the generation units in the high, medium or low voltage level. The exact placement in a grid level is done randomly with the help of the grid generator [81][88][89].

The results of the evaluation of PV systems show that a distinction must be made between roof-mounted and ground-mounted systems. The share of installed capacity of rooftop PV systems in the total installed capacity can be varied depending on the scenario. Typically, the share of rooftop systems is set between 50% and 70%. The rooftop installations are located entirely in the low voltage level. Because 83 % of all ground-mounted systems are connected to the medium-voltage level based on the data analysis, it is assumed that ground-mounted systems are primarily installed in the medium-voltage level. The evaluation of the connection level of wind turbines shows a greater variation. With about 40 % each, the share of wind energy units connected to the medium and high-voltage grid level is almost equal. 13% are directly connected to a high-voltage/medium-voltage substation. This information is used to distribute the wind energy units to specific nodes.

4.1.4 Italy

The main source for (conventional) generation data is from databases collected by RSE during previous projects and collaborations, including the currently on-going monitoring activities of operational power plants (especially hydro technology). Such data included: name of the power plant, region, area, geographical coordinates, fuel type, estimation of the decommissioning year, installed production capacity, storage capacity of pumped hydro power plants.

The collected data set includes the information extracted from the website of Ministero della Transizione Ecologica [56], which is constantly updated with the most recent connection of renewable power plants (RES), including small scale ones. Data from ENTSO-E transparency platform [57] (especially for hydro generation), as well other open sources [58], have been collected as well.

Since the geographical locations of generators are known (even if roughly for some technologies), the matching with the most appropriate electrical grid connection node has been performed according to the Italian grid code [59]:

- Power plants with capacity < 6 MW were assigned to a node of the closest distribution network (the coordinates of primary substation have been used).
- Power plants with capacity 6 MW - 100 MW were connected to (sub)transmission (6 MW - 10 MW) and transmission network (>10 MW), all nodes types with voltage 132 kV - 380 KV.
- Power plants > 100 MW were connected to transmission network, all nodes types with voltage 220 kV - 380 kV.

Finally, a list of operative generators was defined by assigning:

- Generation costs, power capacity and reliability indices to dispatchable generators (thermal based technology).
- Curtailment costs, power capacity and reliability indices to non-dispatchable generators (mostly renewable technology).
- Generation/pumping efficiency, power capacity and reliability indices to hydro pumped storage units.

Power plants with capacity lower or equal than 6 MW were primarily assigned to the closest primary substation. This allocation allowed the construction and testing of the transmission electrical model with the data of 2020 [57].

For the allocation of the new, prospective power plants the current location of the existing connection ENTSO-E results of the Pan- European market analysis to be utilised. In particular, small/medium scale renewables are assumed to be increasing in number of power plants proportionally to the assumed scenario generation capacity. The same principle adopted for allocation of existing power plants on distribution network is used.

4.1.5 Balkan region

The main source for (conventional) generation data is the network model of ENTSO-E for TYNDP 2020. Data from the model include:

- Short name of the power plant.
- Region, area, geographical coordinates.
- Installed production capacity.

Additional generation data were taken from databases collected by EKC during previous projects and collaborations or the information extracted from the ENTSO-E transparency platform [57]. The data included:

- Fuel type.
- Estimation of the decommissioning year.
- Storage capacity of pumped hydro power plants.

Since geographical location of generators is known (even if roughly for some technologies), the matching with the most appropriate electrical grid connection node has been performed according to the following adopted rules:

- Power plants with capacity < 30 MW were assigned to a node of the closest distribution network (the coordinates of primary substation have been used).
- All other power plants were connected to the transmission network, all nodes types with voltage 110 kV - 380 (400) kV with known geographical location.

Finally, a complete data set for operative generators was defined by assigning:

- Generation costs, power capacity and reliability indices to dispatchable generators (thermal based technology).
- Curtailment costs, power capacity and reliability indices to non-dispatchable generators (mostly renewable technology).

- Generation/pumping efficiency, power capacity and reliability indices to hydro pumped storage units.

For the allocation of the new, prospective power plants the current location of the existing connection points will be used for the wind and hydro power plants. For the other power plants the regionalization of the MILES results is utilised. In particular, small/medium scale renewables are assumed to be increasing in number of power plants proportionally to the assumed scenario generation capacity. The same principle adopted for allocation of existing power plants on distribution network is used (see previous section).

4.1.6 Nordic region

For Norway, the grid data set includes partial information about generators and loads connected to these network levels. It represents the situation around 2019 and contains a single load and generation snapshot (power flow solution).

A detailed model of the Norwegian hydropower generation system is associated with the grid model for Norway, including the mapping between the generators and the buses of the grid model.

For Thermal generators:

The 'Matched data' FlexPlan dataset [60] is used as a starting point for thermal generators in Denmark, Sweden and Finland. This source also includes coordinates for each power plant. The assumption is that rather than building new power plants, it is likely that old power plants will be retrofit to use more environmentally friendly fuel types. These data are supplemented by more detailed data in an in-house Excel sheet generator database at SINTEF Energy Research. This data set covers power plants in the Nordic region and some surrounding countries and was last updated around 2018. It also contains information about net electric efficiency and cost related data for the power plants.

For PV and Wind:

For wind power the MILES data is adjusted with datasets of existing wind turbines/farms [61]. For PV the existing capacity is very small for the Nordic region, so for this type the Miles data is used directly.

Power plants with capacity lower or equal than 6 MW (3 MW for Hydro) were primarily assigned to the closest transmission node (primary substation).

For PV, Wind, and Other-RES, installed capacity on the distribution grid is assigned as a function of the total installed generation capacity for the transmission node. This is done by using a piecewise linear curve, which can be different for each country fuel type. For instance, in Sweden and Finland wind in the distribution grid is assigned as follows: All installed capacity up to 5 MW is put on the distribution grid, then 50% of the capacity up to 10 MW, 25% of the capacity up to 20 MW and 0% after that. By doing it this way there will be a difference between distribution grids in wind areas and non-windy areas, while also making sure that no primary substation has more than 10 MW in the connected distribution grids.

Furthermore, to avoid a situation where all distribution grids have some wind/PV/Other-RES, the installed generation capacity is removed and redistributed to the transmission level if it is very low. The cut-off values used for deciding whether to remove and redistribute the capacity in a distribution grid was 4 MW for Wind, 0.1 MW for PV and 1 MW for Other-RES.

For the thermal generators, all 2030 scenarios are the same in the Nordic region. Therefore, a common 2030-database is created by adapting the starting point. This is mainly done by removing old power plants and scaling remaining power plants to achieve total installed capacity equal to scenario values for each fuel type.

For 2040 and 2050 the scenarios only have minor differences, and all changes are reductions relative to the 2030 values. Therefore, updating the dataset for each scenario is done by either downscaling power plants, for minor changes, or removing power plants for larger changes.

4.2 Load

The MILES provides load data, containing the distributed load time-series on higher voltage levels of transmission substations. Similarly, to the workflow required for the identification of generation units, there is the need to perform two activities related to load, which aim to enable the simulation of the foreseen energy scenarios in the six regional cases, including the utilization of demand-side flexibility. These two activities include:

- Identification and characterisation of major loads, which can provide flexibility to the system.
- Creation of a methodology to allocate the regional level load data received from Pan- European market analysis to the required grid nodal level.

The portion of the load belonging to the distribution network has been assigned to each primary substation based on their rated power. As it happens for generation, each load connection to the actual distribution node is based on a pseudo random allocation algorithm, directly managed by the synthetic network generator [47].

In the next sections, these activities are summarized, for each one of the regional cases.

4.2.1 Iberian Peninsula

Two main sources were used for load modelling for the Spanish network. The first was a result of the network analysis while modelling the transmission and sub-transmission network. The open sources (internet) were used to find information regarding load and big consumers. Second, the ENTSO-E scenarios and MILES results were considered.

In the case of Portugal, industrial loads (connected directly to the transmission network) were identified with the collaboration of REN. However, only the type of industry is obtained, as there is no additional information.

The analysis of network nodes led to the identification of several industries, mainly those that were directly connected at a transmission or sub- transmission network. As well others near to substations connected to these networks. Additionally, to the big industries, some other big consumers were found and included in the commercial sector.

The following types of big loads were identified:

- Industrial sector: automotive, cement, chemical (cellulose), gases (air separation), metal (e.g. aluminium), petrochemical, shipyard, steel (the most numerous).
- Commercial sector: airport.
- Transport: high speed train supply.

Types of big loads not directly connected to the transmission network are water treatment, water desalination, paper mill, cement, automotive, gases manufacturing, high speed train.

The general approach for the load connection to the network is next:

- Load data from MILES is assigned to transmission nodes directly.
- If transmission nodes are "load" or "generator" type, the load remains at that voltage level. If they are "substation" type, they are shared to either sub-transmission (if a connection exists at that substation between both networks) or distribution.
- If sub-transmission nodes are "load" or "generator" type, the load remains at that voltage level. If they are "substation" type, the load is shared between the nodes at the distribution network.

In the case of Portugal, the allocation of load to the distribution systems will follow a similar approach of the one explained above but considering as well the information obtained in public reports of the Portuguese DSO [83], where the maximum load values of primary substations are identified as well.

4.2.2 France and Benelux

As in the case of PV and wind generation disaggregation, the demand time series is going to be distributed uniformly, based on the area of influence of the HV substations.

Data on big industries was collected for France and Benelux countries and assign some industrial loads on high voltage levels manually. To do so, also the default demand values of the TYNDP data set are going to be used as reference.

Distributed load will be connected to the network using the approach described before, i.e. uniformly, based on the area of influence of the HV substations (e.g., if there are 5 substations in the area of influence of the substation for which MILES data is provided, each substation will get one fifth of the demand coming from MILES).

4.2.3 Germany, Switzerland and Austria

The major load of each node of the transmission grid is provided by MILES. The load of big industries is subtracted, and the remaining load is distributed to the nodes of the high-voltage grid according to the supply area. It is assumed that this load is generated by underlying medium and low-voltage grids. Accordingly, the load was considered for the distribution grid generation.

The load of big industries was collected from internal ie3 databases which is also used in MILES.

4.2.4 Italy

Different sources of information were used for the collection of data related to loads connected to the Italian network. The time series of the total load is defined per scenario and provided by MILES. Then, it is split into big industries and distribution loads.

Thanks to the availability of RSE internal databases, a list of big loads with geographical information and rated power has been extracted. No real time series for industrial consumption were available and a standard power profile has been used [41], which has been weighted according to the rated power of each industrial load. Loads of big industries have been assigned to the closest transmission network node. In some cases, the voltage level to which the load unit is connected was known and used during the selection of nodes. In other cases, the assignment of the voltage level has been performed based on the rated power.

Then, the total industrial consumption is subtracted from the time series representing the total load and, the resulting one, is assigned to the distribution network and split according to the rated power of the primary substations.

4.2.5 Balkan region

Different sources of information were used for the collection of data related to loads connected to the network of Balkan region. The time series of the total load is defined per scenario and provided by MILES. Then, it is split into big industries and distribution loads.

Thanks to the availability of EKC internal databases, a list of industrial loads with geographical information and rated power has been extracted. Since the number of industrial loads is small (only 4 in Balkan region) and no real time series for industrial consumption were available, loads for all nodes have been kept proportional to the loads given in the initial network model.

The portion of the load belonging to the distribution network has been assigned to each primary substation based on the corresponding load given in the initial network model.

4.2.6 Nordic region

Data about the load distribution for a single snapshot is available with the grid models for Norway and Denmark. For Sweden and Finland, no load data were available.

The best sources found to date for major loads in Norway is a report from NVE about industry from 2013, and [62] and the grid data from [22]. The former has information about approximate localization, approximate annual energy use, and approximate distribution throughout the year. The host page for the report also has some data about expected future energy use for various industries in 2030 and 2040. The grid data from NVE includes connection points for some of the major loads in Norway and Sweden.

5 Environmental indices and input data

The regional case leaders have been asked to provide data of existing thermal power plants that will be still operating in 2030. The template of the requested data can be found in the annex 2. Datasets are briefly described hereafter by region.

In the **Spanish** dataset 30 existing thermal power plants are supposed to be still operating in 2030. These power plants account for 24.7 GW of installed capacity. According to the European Reference Scenario, in 2030 thermal power plant installed capacity in Spain is expected to be 36.9 GW [63]. The fuel sources of these plants are natural gas or biomass (residual) (121 MW). For 18 out of 30 power plants the efficiency is provided, while it can be derived from fuel consumption and annual production for another one. Emission data are complete for 28 plants. No information on the plume rise parameters are provided.

The **Portuguese** dataset comprises 22 plants, for an overall installed capacity of 3.35 GW. For three of these power plants the main activity is power generation (2.9 GW), while the remaining (0.4 GW) are small units employed in the wood and paper industries, whose main activity is to provide heat and energy for the industries' self-consumption. For the first three the fuel source is natural gas, for the others is biomass. For biomass the source is either fuelwood or wood residues or by-products. There is no information on the subdivision between primary and residual biomass. The installed capacity of thermal power plants in Portugal in 2030 according to the European reference Scenario is 5.75 GW [63].

The **France and Benelux** dataset is filled with 131 power plants (22 of them nuclear) with a total installed capacity of 118.9 GW. 56 plants are in France, with a total installed capacity of 81.8 GW (of which 63 GW are from nuclear power plants). 22 plants are in Belgium, with a total installed capacity of 13.0 GW. No power plants are operating in Luxembourg. 53 power plants are in the Netherlands, with an installed capacity of 24.1 GW. 57 plants are characterized by their efficiency; for the other plants the efficiency cannot be derived from fuel consumed and electricity produced, because of insufficient data. Greenhouse gas emission data are not provided but for 3 plants. For air quality data, SO2, NOx and PM10 annual emissions are provided for 33 plants. No information is provided regarding plume rise parameters.

For **Germany, Austria, Switzerland**, there are 38 plants in Germany (installed capacity 20.3 GW); 21 in Austria (5.1 GW) and 3 in Switzerland (0.132 GW). The dataset thus comprises 62 plants, which represent a total installed capacity of 25.5 GW. 48 of the 62 plants are described with plant efficiency. CO_2 annual emissions are provided for 60 plants, as well as N_2O annual emissions for 4 plants. No CH_4 data are available. For the air quality part, 25 plants present CO_2 (or SOx) data; 34 NOx; 1 VOC; 4 CO; 1 NH_3; 16 PM10 (all expressed in terms of tons/year). For 13 plants information about the number of stacks and the height are provided.

The **Italian** regional case includes 117 fossil fuel and 1565 biomass fuel power plants, respectively accounting for 49.1 GW and 5 GW of installed capacity.

Emission data for fossil fuel plants have been collected from the sources listed in the following section, "Air quality" paragraph. The availability of data is as follows:

- 94 plants included in the LCP reporting.
- 91 plants included in the E-PRTR.
- 66 plants for which EMAS declarations have been found.

No data was found for 13 plants, totalling 0.47 GW installed capacity.

Only 20 among the biomass fuelled plants are included in the available data sources, therefore most of the emissions will be estimated based on installed capacity, using well-established emission factors.

The **Balkan region** dataset encompasses 35 power plants (12.3 GW) expected to be operating in 2030. This list includes both running plants and plants not yet in operation. These plants are fuelled with natural gas, or hard coal or lignite. 10 plants are in Serbia (6.0 GW); 5 in Croatia (1.5 GW); 4 in Slovenia (1.1 GW); 7 in Bosnia Herzegovina (1.9 GW); 4 in North Macedonia (0.9 GW), 2 in Montenegro (0.5 GW); 3 in Albania (0.3 GW). For every power plant efficiency and CO_2, SO_2, NOx and PM2.5 emission factors are provided. PM10 annual emissions are available for 20 power plants. For 16 plants the following plume rise parameters are provided: number of stacks, stack height and diameter.

The dataset of the **Nordic region** encompasses 66 power plants in Denmark, Finland and Sweden. In Norway in 2030 no thermal power plants are expected to be operating. 41 of these power plants are in Finland (7.2 GW); 15 are in Denmark (5.4 GW); 10 are in Sweden (3.6 GW), for an overall installed capacity of 16.2 GW in the Nordic region. With four exceptions in Finland, all the plants are multi-fuel. For each power plant, efficiency is provided. For the carbon footprint section, 22 plants present CO_2 emission data, 2 have also CH_4 data, no data on N_2O are provided. For the air quality section, all the power plants are filled with SOx, NOx and PM10 emission data and the reference yearly profile. Plume rise parameters are provided fully for 8 plants, partially for 12. No data on nuclear plants are provided for Finland.

5.1 Carbon footprint data

Besides the Spanish case, all the datasets have gaps to some extent, and those need to be filled before performing the carbon footprint analysis. To do so, the following two approaches were taken:

For datasets providing the required technical data (efficiency or fuel consumption and energy production) but lacking all or part of the emission data, the emission factors of the IPCC have been considered [64]. These emission factors provide the amount of CO_2, CH_4 and N_2O for a unit of fuel (MJ) burned in the power plants. Combined with the power plant efficiency (or fuel consumption and electricity generated at the power plant) it is possible to derive the direct CO_2 equivalent of the kWh generated at the power plant.

For plants with insufficient technical data (e.g. efficiency is not available or it cannot be calculated) the average efficiency of plants with the same technology has been used.

As far as biomass fuel power plants are concerned, where there is no distinction between the biomass origin (residual vs not residual) the current ratio between biomass and residual biomass in the Eurostat dataset is applied [65].

5.2 Air quality

In all regional cases, emissions or emission factors for air quality related pollutants are lacking to some extent. The main sources of data used to fill the gaps are:

- the European and national Pollutant Release and Transfer Register (PRTR) databases [66], if not already used by the representatives of the region.
- reported data on large combustion plants (LCP) covered by the Industrial Emissions Directive – 2010/75/EU, Article 72 [67].
- European Eco-management and Audit Scheme (EMAS) declarations [68] (European Commission, 2011) for those plants that voluntarily chose to adopt the scheme.
- The EMEP/EEA air pollutant emission inventory guidebook 2019 [69].

Missing parameters to be used for plume rise estimation are filled using values typical of power plants of the same technology and size.

5.3 Environmental models

The simplified model used to evaluate air quality impact is derived through a Taylor expansion approach based on a full 3D chemistry and transport model (CTM) able to reproduce all chemical and physical processes which air pollutants undergo in the atmosphere. The CAMx model (Comprehensive Air Quality model with extensions, [70]) is applied together with the embedded DDM algorithm (Decoupled Direct Method for sensitivity analysis in a three-dimensional air quality model, [71]). In addition to simulating concentration fields over domains covering all the regional cases, CAMx/DDM estimates sensitivity of those fields to variations in the emissions associated with the thermoelectric power plants considered in the project. This allows to link variations in production of each plant to variations in emissions, and consequently in pollutant concentrations over the fallout areas. Finally, health impact on population living in those areas, as well as the related costs, are derived from concentration variations. For a detailed description, see Section 6 deliverable D 1.2 of the FlexPlan project [4].

The modelling grids used for the regional cases runs of the CTM have been chosen in order to balance coverage and resolution of the output fields on the one hand, and computational efficiency on the other, given the complexity of the involved numerical code. As the Nordic countries of Nordic regional case is located peripherally on the northeast with respect to the others, a dedicated domain is used, while all the other regional cases are covered by the same single domain. The two grids are defined in the WGS84 geographic Coordinates Reference System (CRS) as detailed in Table 5-1, with a grid spacing of 0.125° (1/8 degree). This choice was made considering the input data CRS, which is the same for meteorological and emission files (see below). An overview of the two grids is shown in Figure 5-1 and Figure 5-2.

Grid	min. longitude	max. longitude	min. latitude	max. latitude	No. of cells along x	No. of cells along y	Total no. of cells
Nordic RC	6	32	54	67	208	104	21632
Other RCs	-11	24	34	56	280	176	49280

Table 5-1 Grid parameters for CAMX/DDM simulations

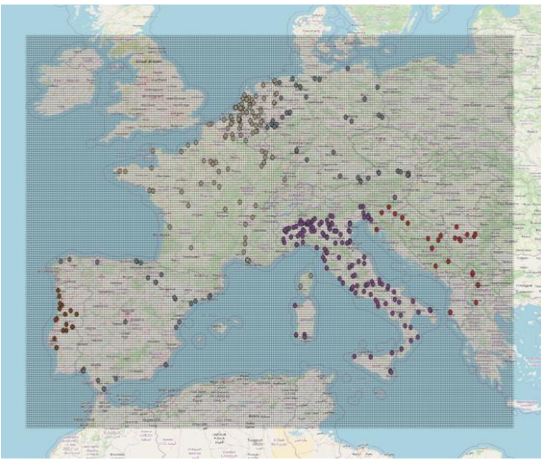

Figure 5-1 - CAMX/DDM modelling grid for all regional cases but the Nordic countries regional case

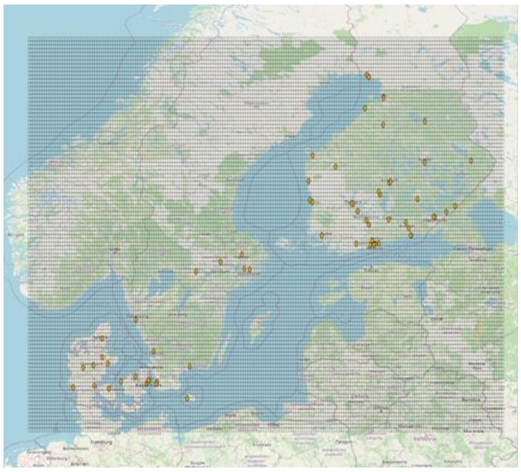

Figure 5-2 - CAMX/DDM modelling grid for the Nordic countries regional case

Meteorological fields used for driving advection and dispersion in the photochemical model are taken from the ERA5 dataset (provides hourly estimates of a large number of atmospheric, land and oceanic climate variables), made available by the ECMWF (European Centre for Medium-Range Weather Forecasts) [73]: a comprehensive reanalysis of forecasts performed by the IFS model and data assimilation system [74], which assimilates as many observations as possible in the upper air and near surface. Data can be retrieved on a regular latitude-longitude grid, with a resolution of up to 0.25°. ERA5 fields are then interpolated on the higher-resolution grids and further parameter variables needed by CAMx are computed from the available ones.

Pollutant emissions from the investigated power plants are assigned to point sources in the relevant grid cell based on their geographical coordinates. Vertical position and air mass buoyancy characteristics are calculated using the ancillary data (stack height and diameter, flue gas exit flow rate, temperature and velocity) received from the RCL, or estimated if missing.

Besides this critical input, further data are needed to describe the overall emissions forcing the concentration field on the modelling domain. The CAMS_REG_AP emission inventory produced by TNO [75] will be used, processing the data with the HERMES emission modelling framework [76]. Total annual data provided for each pollutant and each activity sector on 0.0625° x 0.125° lat-lon cells is then spatially, chemically and temporally disaggregated to obtain hourly emissions of the relevant chemical species on each grid cell of the target modelling domain.

In order to correctly simulate the chemical equilibrium inside the modelling domain, boundary conditions will be derived from concentration fields produced by large scale chemical model runs, available from European services (e.g. CAMS).

A winter and a summer month of the year 2017 will be simulated with a reference load profile assigned to power plants in each regional case, derived from available data. For each regional case, the methodology already used for the Italian test case will be followed to obtain the Imp_{ref} and $Prod_{ref}$ coefficients as defined in the deliverable D1.2 of the FlexPlan project [1]. In brief, an estimation of the sensitivity of concentration values to emission changes (i.e. production changes) for each grid cell, for each hour of the day and for each season is calculated as an average over one month's worth of CAMx simulation data, where the investigated sources are those associated to power plants. This way, daily and seasonal variability in dynamics and chemical evolution of the air masses in the atmosphere is considered.

Finally, the effect of air pollution on health is calculated combining variations in pollutant concentration for each model grid cell (estimated annual average PM2.5 concentration), municipality-level population data and baseline incidence rates with a log-linear model. Yearly mean concentration of pollutants is calculated with CAMx model, population data comes from EUROSTAT population dataset [76] and all the incidence rates come from HRAPIE (Health risks of air pollution in Europe) project [77]. As indicated in the latter report, we assume that premature mortality affects only adults 30 years of age or older, and we use a value of relative risk (RR) per 10 µg/m3 equal to 1.062 (1.040-1.083, 95% confidence interval) for premature mortality. In order to allow a more robust economical evaluation of the premature mortality impact, health impacts are expressed as Years Of Life Lost (YOLL).

6 Costs and economic data

The simulation of the full planning procedure to be performed in each one of the regional cases requires the existence of fuel costs and other economic data necessary to characterize all involved flexibility elements.

As regarding generation costs, there is the need to characterize the fuel costs. This is required for the execution of the OPF (that allows to identify grid congestions and reinforcement needs) of the different scenarios considered in FlexPlan. Due to the lack of complete information regarding generation costs from publicly available sources, it was decided to use ENTSO-E commodity prices from the TYNDP 2020 in the regional case studies. Another assumption performed includes the linearization of generation cost curves, as there is no possibility to obtain real cost curves for the different thermal generators existing at pan-European level (and required for the different regional cases). Table 6-1 contains the different fuel costs used in the scope of FlexPlan. These are maintained the same as present in TYNDP 2020 from ENTSO-E.

	Year	2030 DE	2040 DE	2050 DE	2030 GA	2040 GA	2050 GA	2030 NT	2040 NT	2050 NT
Euro per megawatt	Nuclear	1.7	1.7	1.7	1.7	1.7	1.7	1.7	1.7	1.7
	Lignite	3.96	3.96	8.3	3.96	3.96	3.96	3.96	3.96	3.96
	Hard coal	15.48	24.9	34.3	15.48	24.88	34.27	15.48	24.876	34.27
	Gas	24.88	26.3	27.7	24.88	26.32	27.76	24.88	26.32	27.76
	Oil	0	0	0	0	0	0	0	0	0
	Light oil	73.8	79.92	86.04	73.8	79.92	86.04	73.8	79.92	86.04
	Heavy oil	52.56	61.92	71.28	52.56	61.92	71.28	52.56	61.92	71.28
	Oil shale	8.28	8.28	8.28	8.28	8.28	8.28	8.28	8.28	8.28
	Diesel	37.31								

Table 6-1 Fuel prices

In addition to the fuel costs, there are additional economic data required. This includes data allowing to have an economical characterization of flexibility measures and grid expansion measures. All grid expansion candidates (including flexibility elements such as battery storage) will be fully characterized and a CBA is performed, as defined in the scope of work on the planning tool and flexibility elements identification and characterization. Regarding the economic characterization of flexibility solutions, these can be mainly summarized as:

- Generation flexibility – flexibility from the generation side is achieved through the definition of curtailment costs. These will be established taking into consideration the different technologies and available data sources.

- Demand side flexibility – flexibility from the demand side is characterized using different technical and economic parameters usually used in demand side flexibility characterization. The Value of Loss Load (VOLL) is used to characterize involuntary load curtailment due to contingencies (outages) and power supply interruptions, and should be distinguished from other forms of (flexible) demand reduction. The parameter can be assigned to all loads and not just flexible loads. Regarding voluntary load reduction or shifting, different parameters are considered, including: i) compensations for decreasing or shifting power consumption, allowing to compensate for voluntary demand reduction or shifting of flexible loads, respectively; ii) definition of maximum energy not consumed, allowing to establish a maximum yearly value of voluntary load reduction, iii) recovery periods, establishing maximum periods in which a mobilized flexibility action (e.g. load reduction) needs to be compensated and iv)bounds for upward or downward shift compensations.

7 Common post processing of the market results

Data received from both, the market simulations and the regionalization process performed in WP4 was put available to WP5 in a regional level. However, in order to enable the simulation of regional cases, a full nodal description is necessary, in order to create the corresponding time-series and hourly scenarios to be simulated. In this section, a short description of the adaptation approach followed by each regional is provided. It is important to note that the full understanding of the implemented methodologies also takes into consideration the generation and load adaptation methods, presented in section 4.

7.1 Used approach for adaptation of Pan-European market results

MILES results for market simulations include as well cross border power flows, one for each neighbouring country. The general approach to be followed by all regional cases is that for each border (e.g. France-Germany) the power profile is divided proportionally to the actual capacity of each connection segment. Additionally, MILES results also included the total generation provided by dispatchable units and hydro pumped storage. These profiles have been used for the verification of the consistency of the scenarios, but are not part of the data set requested by the planning tool as the dispatching is performed in the toolchain of FlexPlan.

In the **Iberian regional case**, a slightly different approach is used for Spain and Portugal, regarding the adaptation of pan-European market results from WP4. For Spain, the MILES information has been used as reference for non-dispatchable generation units and for loads. Dispatchable generation and biomass (here considered non-dispatchable following the MILES approach) have been identified at network model level. The matching between the data in MILES and the information retrieved on the current network was done, so that they were coherent. As non-dispatchable information comes from MILES, this was made for dispatchable generation including thermal and hydro.

For Portugal, as additional data is available for both dispatchable and non-dispatchable power plants (from REN), values from MILES are adapted accordingly. This means that there is an intermediary step, corresponding to a change of location in some installed capacities. This step only implies a change of geographical location and overall installed capacity and national generation time-series are kept. This step is performed with the collaboration of REN, and aims at achieving a realistic generation profile (considering the expected geographic distribution of the different generation technologies). The surplus of installed capacity existing in MILES (accounted for as new facilities to be deployed until 2030) was not changed and we assume MILES outputs are correct. These are applicable mostly to wind and PV.

The information from MILES was received at transmission node level. This was disaggregated to the sub-transmission level with a slightly different approach for generation and loads:

- Generation: sub-transmission nodes were assigned to transmission nodes considering proximity (similar solar, wind conditions).
- Loads: sub-transmission nodes were assigned to transmission nodes considering connectivity (connection between transmission and sub-transmission at certain nodes).

The selection of the generation and load that had to be connected to distribution, has been performed in different ways depending on the technology, as explained in sections 4.1.1 and 4.2.1.

For **France and Benelux**, once the transmission, sub-transmission and distribution networks are fully available, including the location of generators, loads and storage units, the profiles returned by MILES will be disaggregated which is providing the time series for total consumption for each considered geographical location. One additional disaggregation will be on distributing the cross-border power flows as provided by MILES over the different tie-lines present in the system.

For **Germany, Switzerland and Austria**, the assumptions mainly include the connection points of the distribution grid, which was not covered in MILES. These assumptions are presented in 4.1.3.

For **Italy**, once the Italian transmission and distribution networks were available, including the location of generators, loads and storage units, the profiles returned by MILES have been split according to:

- The geographical area which, for the Italian case, has been selected to coincide with provinces and big city neighbouring countries.
- The rated power of generators, loads and storage units of the reference case. Capacity upgrades, which are dependent on the technology, are considered in this process.

MILES has provided a time series for total consumption for each considered geographical location. Thus, for each area:

- The big industries consumption has been subtracted from the total profile.
- The resulting profile has been divided according to the number of primary substations included within the area.
- The split has been performed considering the rated power of the primary substations, provided by the reference case (2020).

Looking at the power profiles of non-dispatchable generation, the same geographical areas have been considered. For each area and depending on the technology:

- The rated power or the amount of power plants was revised in order to match the scenario data.
  - For some kinds of technology (e.g. wind, hydro), the existing power plants in the area were scaled according to the scenario data.
  - For other technologies (e.g. photovoltaics), the amount of power plants was redefined according to the scenario data.
- The production profile has been split according to the amount and rated power of generators belonging to the considered area.

For the **Balkan region**, MILES time series of consumption for each considered geographical location were used. These time-series have been summed up per each country and after that:

- Loads for all primary stations have been kept proportional to the loads given in initial network model.

Looking at the power profiles of non-dispatchable generation, the same geographical areas have been considered. For country and depending on the technology:

- The rated powers of existing wind, hydro and solar power plants were scaled according to the scenario data.
- The production profile has been split according to the amount and rated power of generators belonging to the considered country.

For the **Nordic region**, for most cases where real world data was not available, the MILES profiles have been used distributed to the network busses and used directly. A python script has been created which compares the coordinates associated with MILES profiles with the coordinates of the power network buses. First each of the network busses are associated with the closest MILES profile. Then, the MILES profiles that have not been associated with any network busses are associated with the closest grid bus. Now, all network busses and MILES profiles should be associated with at least one of the other. Finally, each MILES profile is divided evenly among all its associated network busses.

Where real life data is available, which is the case for wind in Denmark [90] and Norway, that data has been used to adapt the MILES profiles. The capacity in each network bus is found by connecting the real-world generators to the closest network bus. Since the MILES profiles represent the wind conditions at each physical location, they are used on a per-unit basis for the new capacities. The profiles are then adjusted to maintain the same production per hour for each region.

8 Common pre-processing procedures for the case studies

8.1 ICT aspects and JSON file structure

The workflow of the software developed under the FlexPlan project involves interaction between different modules, each one led by a different developer. Thus, it needs to be (a) customizable, (b) easy to understand, and (c) lightweight. An overview of the different information that is needed to be transferred between the users, the optimization engine (WP 3) and the assessment of potential flexibility (WP 2) is depicted in Figure 8-1.

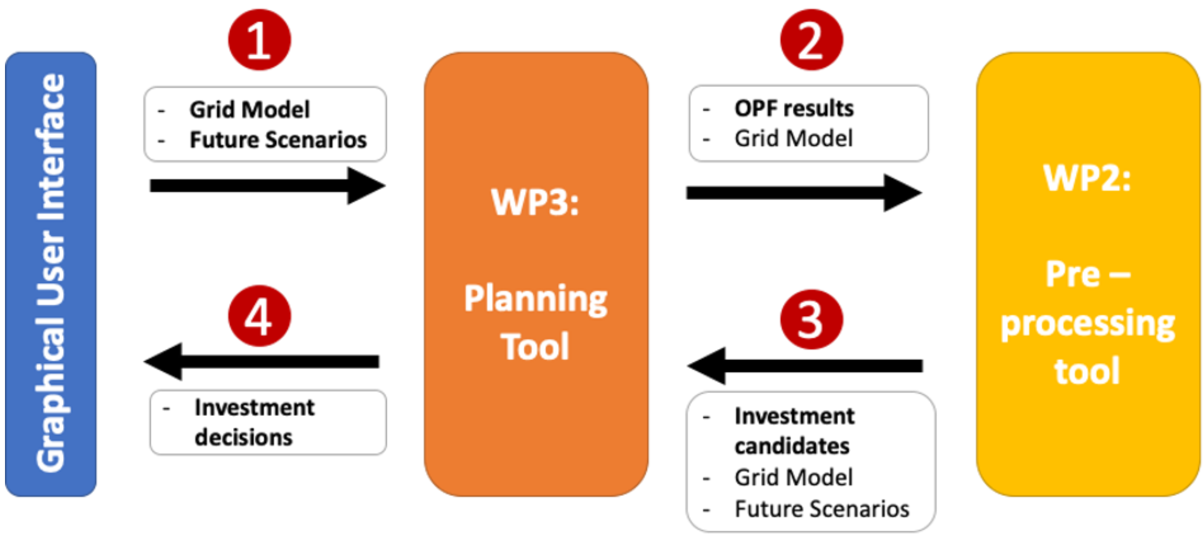

Figure 8-1 - The workflow between Planning Tool and Pre-Processing tool

The chosen format for the exchange of information is JSON, which is a common practice by both industry and research. Among other properties, JSON allows to represent objects in many programming languages as human-readable text that can be easily exchanged through APIs and databases. Also, it is easy for machines to parse and generate, which reduces the execution time of the tool. Furthermore, it is supported by Python natively (language of choice for the implementation of the planning tool, both for mathematical modelling or results generation, as well as for internal and external communication through several API layers), meaning the expected performance when reading the files is high, as well as numerous useful built-in functions are available.

The JSON files which will collect the data and models obtained from WP 5 need to be aligned with the needed input for the optimization problem designed in the deliverable D1.2 of the FlexPlan project [1]. In such a document, one can consult Section 2, which details the sets, entities and indices used in the planning tool, together with the optimization variables and parameters.

8.2 JSON file structure

The following Table 8-1 – Table 8-4 provide detail of the content of each of the files mentioned in Figure 8-1.

Data	Details
Grid model	• The topology of the power system, providing information related to buses, lines, converters, transformers, loads, generators, and storage devices. • Here is included equipment technical characteristics, i.e. static values which do not vary through the year (eg. generators's nominal capacity or branches' thermal rating). • Similar to equipment profile (EQ) data files, used in the Common Grid Model Exchange Standard (CGMES) by ENTSO-E [3].
Future scenarios (2030/2040/2050)	• Data related to buses, loads, generators, and energy storage devices. • Here it is included data which is time-dependant through a year (eg. load and generation profiles). • Similar to the steady-state hypothesis (SSH) data files, used in CGMES.

Table 8-1 - Input files for the planning tool

Data	Details
Optimal Power Flow (pre-investment)	• Outcomes of the non-expanded optimal power flow for years 2030, 2040 and 2050. • Expressed as Locational Marginal Prices (LMPs) at the buses, power flow directions through branches, transformers and converters and Lagrange Multipliers (LMs) associated with each branch. Additionally, the PTDF matrix of the network is provided.
Grid model	• Same as Input 1

Table 8-2 - Output files of the planning tool after OPF simulation

Data	Details
Investment candidates	• Characteristics of the investment candidates • Expressed as a combination of resource location, type of flexibility technology candidate, alongside with its size and cost
Grid model	• Same as Input 1
Future scenarios	• Same as Input 1

Table 8-3 - Input files for running a stand-alone network expansion planning problem

Data	Details
Investment decisions	• List of the optimal investment decisions based on the provided candidates expressed as Boolean variables

Table 8-4 - Output files for running an expansion planning problem

Data	Details
Generators	• The list of all generators with their type (PV or wind), their connected bus and their generated active power variants
Loads	• The list of all loads with their connected bus and their demand reference variants
Granularity	• The length of the variants after scenario reduction (daily, weekly, monthly or yearly)
Number of clusters	• The number of variants to select

Table 8-5 – Input files for running the scenario reduction process

Data	Details
Scenario probabilities	• The list of selected variants together with their probability

Table 8-6 - Output files after running the scenario reduction process

8.3 Used approach for creation network models in JSON format

Within the FlexPlan project, three visions (scenarios) for each target year (2030, 2040, 2050) based on diverse storylines, resulting in a total number of nine pan-EU scenarios have been defined.

To determine the optimal grid expansion on a nodal level, the pan EU-scenarios have been disaggregated to individual nodes in the transmission and distribution grids such that they can be used by the new planning tool. Furthermore, the dependency of non-dispatchable units, e.g. wind, solar and partly hydro power generators, on local climatic conditions was considered to obtain realistic results within the regional studies. For the consideration of uncertain input variables, such as intermittent renewable generation, the databases were set up, which contain historic weather conditions, i.e. a time series of historic wind speed near the plants' locations. The databases incorporate climatic conditions (in terms of hourly capacity factors) for 40 years (1980 – 2019). The recording of one historic year consisting of 8,760 hours and it is known, that the consideration of only one historic year is not representative for future years, as climatic conditions can vary considerably from one year to another. As the amount of generated time series data, on nodal level for uncertain input parameters becomes large, especially due to the consideration of various climatic conditions (variations), a scenario reduction approach has been applied to reduce the complexity. The purpose of the scenario reduction is to choose a representative but reduced set of time series out of the 40 generated meteorological variations. Of the 40, there are 35 years with data on demand, PV, and wind, all correlated on time. Those 35 correlated years are used as input for the scenario reduction. More information on scenario reduction can be found in deliverable D1.1 of the FlexPlan project [82].

JSON files have been created by using code, written in different programming languages, for converting the regional networks and market results data from its modelling tools into required JSON format, which are input files for the new planning tool. The two main input files were created for each region:

- JSON file with regional network data (TN and DN) for optimal power flow calculation.
- JSON file with 35 meteorological variations of load, wind, and PV production for scenario reduction.

As it was decided to use five meteorological variations, for each scenario the six regional networks with five meteorological variations each were created and tested by a new planning tool. Codes for generating JSON data were written by representatives of each region, here below in Table 8- is an overview given of the modelling tools used by region and programming languages used for writing the program for the data adaptation.

Region	Existing modelling tool	Programming languages
Iberian Peninsula	PSS/E, DigSILENT Power Factory	Python
France and Benelux	PowerModels.jl	Julia
Germany, Switzerland, Austria	MATPOWER	Python
Italy	MATPOWER	MATLAB
Balkan region	PSS/E	Excel/Python
Northern countries	pandapower	Python

Table 8-7 - Overview of the modelling tools and programming languages.

9 Conclusions

This deliverable has clarified the modalities adopted for assembling the input data needed for the run of the six FlexPlan regional cases. As explained, these modalities differ from regional case to regional case due to the availability of different kinds of data.

Of course, the procedure of establishing the data set are delicate and affected by uncertainty. Guesses and hypotheses are sometimes needed whenever primary data sources are not available (that is for instance true for sub-transmission and distribution networks). That means that the analysis of the results of the simulations will require particular attention and critical sense to understand if the obtained results are in line with expectancies, if there are some strange outcomes that force a reflection on some input data, etc.

It is very likely that the set of input data will be at least partially revised after the first run of the simulation. The data set will be iteratively refined based on experiences we develop by running increasingly to arrive to the final data set.

Particular attention will be devoted to distribution networks that, due to their big extension, would require to set a number of input data that, even with the assumption to simulate only the highest voltage level, i.e, the one connected to primary substations, would turn out to be too much for ensuring computational tractability. Additionally, no public source could be used to retrieve such data. Therefore, it was decided to set up “synthetic networks” replicating the macro-structure of the real ones, yet in a simplified way and whose parameter are calculated on the basis a set of statistics drawn from real networks. Also, for that, adjustments will be possible after testing the results of the regional cases.

A final point of attention is constituted by the linearized parameters assumed to internalize the air-quality aspects, calculated on by means of simulation carried out on complex 3D environmental models. This kind of simplification, yet allowing to retain the linearity of the overall model, is something new that is tested for the first time in the FlexPlan project. Thus, the results have to be attentively scrutinized and adjustments to the procedure and on the input data are possible as well.

10 Annex 1 – Network modelling assumptions summary

Regional Case	Country	Transmission and sub-transmission data source	Nodes, nr	kV	Modelling tool	Grig data (transformers and power lines) parameters	Source for localization of the grid nodes
Iberian Peninsula	Spain,	ENTSO-E model, TSO grid map for sub-transmission	1846	110 – 132; 220 – 400	PSS/E	Standard electrical parameters for sub-transmission network for Spain and ENTSO-E model for transmission network	Open sources (OSM), Google Maps
	Portugal	ENTSO-E model		150 - 400		ENTSO-E model	TSO data
France and Benelux	France	OpenStreetMap, ENTSO-E model	12209	63 - 380	PowerModels.jl	Standard electrical parameters for sub-transmission network for France and ENTSO-E model data for transmission networks	Open sources (OSM), Google Maps
	Netherlands	ENTSO-E model		50 - 380			
	Luxembourg			13.7 - 220			
	Belgium			70 - 380			
Germany, Switzerland, Austria	Germany	ENTSO-E model, OpenStreetMap	6250	110 - 380	MILES, PowerFactory	ENTSO-E model, MILES grid data	ENTSO-E Map, MILES-database MTE, OpenStreetMap
	Austria						
	Switzerland						
Italy	Italy	MTE website[19], OpenStreetMap	4180	132 - 380	MATPOWER	Standard electrical parameters	

Regional Case	Country	Transmission and sub- transmission data source	Nodes, nr	kV	Modelling tool	Grig data (transformers and power lines) parameters	Source for localization of the grid nodes
Balkan region	Serbia, Macedonia, Albania, Montenegro, Bosnien, Herzogovina, Croatia, Slovenia	ENTSO-E model	1993	110 - 400	PSS/E	ENTSO-E model	EKC's internal database, open sources
Nordic region	Norway	TN under a Non-Disclosure Agreement	3200	132 - 420 10 - 66	pandapower	Provided with model	Open sources
	Sweden	OpenStreetMap, PyPSA-EUR model		110 – 132 220 – 400		Standard electrical parameters	
	Finland	OpenStreetMap, PyPSA-EUR model		110 – 132 220 – 400		Standard electrical parameters	
	Denmark	Danish TSO		132 – 400		Provided with model	

11 Annex 2 – Environmental data template

General data								Carbon footprint data									Air quality data																							
		Plant location									total annual emissions [tons/year]			emission factor [kg/MJ]			total annual emissions [tons/year]								pollutant emission factor [kg/MJ]									plume rise calculation parameters (stack data)						
Plant Id	Plant Name	latitude	longitude	fuel type	Technology	installed capacity [MW]	efficiency (%)	type of biomass (for biomass plants) % of uranium enriched and % of MOX used (for nuclear plant)	Fuel Consumption (MJ/year)	total annual production (in reference year)(kWh)	CO_2	CH_4	N_2O	CO_2	CH_4	N_2O	SO2	NOx	VOC	CO	NH3	PM10	PM2.5	total	SO2	NOx	VOC	CO	NH3	PM10	PM2.5	total	reference yearly profile (attachment)	number of stacks	height [m]	diameter [m]	flow rate [m3/s]	exit velocity [m/s]	exit temperature [K]	emission factor (emissions by fuel consumption) [kg/ton]

12 References

[1] FlexPlan Consortium, “D4.1 Pan-European scenario data” 2020.

[2] FlexPlan Consortium, “D4.2 Pan-European scenario data”, 2021.

[3] ENTSO-E data, 2nd ENTSO-E Guideline, "For Cost Benefit Analysis of Grid Development Projects", 2018.

[4] FlexPlan Consortium, “D1.2 Probabilistic optimization of T&D systems planning with high grid flexibility and its scalability” 2020.

[5] e-HIGHWAY 2050 “D 6.3 Modular Development Plan of the Pan-European Transmission System 2050”, December 2015.

[6] “Sistema Eléctrico Ibérico”, Red Eléctrica de España, [Online]. Available: www.ree.es, last accessed: 17/06/2021] January 2018.

[7] OSM, [Online]. Available: https://www.openstreetmap.org/#map=5/42.088/12.564, last accessed: 17/06/2021].

[8] “Sistema Eléctrico Peninsular”, Red Eléctrica de España, [Online]. Available: www.ree.es, last accessed: 17/06/2021], December 2017.

[9] PyPSA-EUR (GridKit extraction from ENTSO-E-E map) [Online]. Available: https://github.com/PyPSA/pypsa-eur, last accessed: 17/06/2021].

[10] Reseaux d'energies, "lignes-aeriennes-rte.". [Online]. Available: https://opendata.reseaux-energies.fr/explore/dataset/lignes-aeriennes. rte/information/?disjunctive.etat&disjunctive.tension Last accessed: Thursday, July 1, 2021].

[11] Reseaux d'energies, "Lignes souterraines RTE.". [Online]. Available: https://opendata.reseaux-energies.fr/explore/dataset/lignes-souterraines-rte/information/?disjunctive.etat&disjunctive.tension. Last accessed: Thursday, July 1, 2021].

[12] Reseaux d'energies, " Enceintes de poste RTE.". [Online]. Available: https://opendata.reseaux-energies.fr/explore/dataset/enceintes-de-poste-rte/information/?disjunctive.etat&disjunctive.tension Last accessed: Thursday, July 1, 2021].

[13] Coffrin, Carleton, et al. "Powermodels. jl: An open-source framework for exploring power flow formulations." 2018 Power Systems Computation Conference (PSCC). IEEE,. [Online]. Available: https://lanl-ansi.github.io/PowerModels.jl/stable/] 2018.

[14] ENTSO-E Map, [Online]. Available: https://www.entsoe.eu/data/map].

[15] German Grid Development Plan and List of Projects: [online: https://www.netzentwicklungsplan.de/de].

[16] TYNDP: [Online]. Available: https://tyndp.entsoe.eu/].

[17] Homepage of Open Street Map Germany: [Online]. Available: https://www.openstreetmap.de/].

[18] J. Kays, A. Seack, T. Smirek, F. Westkamp, and C. Rehtanz, “The Generation of Distribution Grid Models on the Basis of Public Available Data”, IEEE TRANSACTIONS ON POWER SYSTEMS, VOL. 32, NO. 3, MAY 2017.

[19] Ministero della Transizione Ecologica [Online]. Available: http://sinva.minambiente.it/mapviewer/index.html?collection=http://sinva.minambiente.it/WMC/Collection/VA/D5DB339E-DB14-9144-B043-B141DCABC108&v=full].

[20] ENTSO-E: [Online]. Available: https://www.entsoe.eu/data/map/].

[21] MATPOWER, [Online]. Available: https://matpower.org/docs/MATPOWER-manual-7.1.pdf].

[22] Norges vassdrags- og energidirektorat (NVE) [English: "The Norwegian Water Resources and Energy Directorate", i.e. the Norwegian energy regulator]. Web site: https://www.nve.no/english/. (2020a; the data set is not publicly available.)

[23] Energinet, "Transmission system data". [Online]. Available: https://en.energinet.dk/Electricity/Energy-data/System-data (accessed Jan. 19, 2021).

[24] J. Hörsch, F. Hofmann, D. Schlachtberger, and T. Brown, 'PyPSA-Eur: An open optimization model of the European transmission system', Energy Strategy Rev., vol. 22, pp. 207–215, Nov. 2018, doi: 10.1016/j.esr.2018.08.012.

[25] G. Boeing, 'OSMnx: New methods for acquiring, constructing, analysing, and visualizing complex street networks', Comput. Environ. Urban Syst., vol. 65, pp. 126–139, Sep. 2017, doi: 10.1016/j.compenvurbsys.2017.05.004.

[26] G. Boeing, 'Urban Street Network Analysis in a Computational Notebook', REGION, vol. 6, no. 3, Art. no. 3, 2019, doi: 10.18335/region.v6i3.278.

[27] Norges vassdrags- og energidirektorat (NVE), "NVE Atlas – NVE Kraftsystem Nettanlegg". [Online]. Available: https://atlas.nve.no/Html5Viewer/index.html?viewer=nveatlas# (accessed Nov. 16, 2020b).

[28] SmartNet Consortium: " D4.2 Scenario setup and simulation results". [Online]. Available: http://smartnet-project.eu/wp-content/uploads/2019/07/D4.2_20190628_V1.0..pdf], 2019.

[29] EDP Distribuição, "Caracterização das redes de distribuição a 31 Dezembro 2019,", [Online]. Available: https://www.e-redes.pt/pt-pt/caracterizacao-das-redes-de-distribuicao-2019] 2020.

[30] European Commission Joint Research Center, DiNeMo (Distribution Network Models), [Online]. Available: https://ses.jrc.ec.europa.eu/dinemo. Last accessed: Thursday, July 1, 2021

[31] Grzanic, M., Flammini, M. and Prettico, G., Distribution Network Model platform: a first case study, ENERGIES, ISSN 1996-1073 (online), 12 (21), 2019, JRC118032.

[32] S. Meinecke, D. Sarajlić, S. R. Drauz, A. Klettke, L.-P. Lauven, C. Rehtanz, et al., "SimBench—A Benchmark Dataset of Electric Power Systems to Compare Innovative Solutions Based on Power Flow Analysis," Energies, vol. 13, no. 12, p. 3290, Jun. 2020.

[33] C. Kittl, D. Sarajlić, C. Rehtanz, "k-means based identification of common supply tasks for low voltage grids", 2018 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe), 2018.

[34] A. Seack, J. Kays, and C. Rehtanz, "Generating low voltage grids on the basis of public available data", CIRED Workshop 2014, Rome, 2014.

[35] Simbench Documentation, [Online]. Available: https://simbench.de/wp-content/uploads/2019/08/simbench_documentation_de.pdf.

[36] Information on German Municipalities, [Online]. Available: https://www.destatis.de/DE/Themen/Laender-Regionen/Regionales/Gemeindeverzeichnis/_inhalt.html;jsessionid=CC3F5BE0CB488F1F9F3017EAB74A5018.live712.

[37] Development plans released by distribution system operators [DSO websites https://www.e-distribuzione.it/, [Online]. Available: https://www.areti.it/home.php, https://www.unareti.it/unr/cms/unareti/, https://www.ireti.it/energia-elettrica, https://www.set.tn.it/content/home, http://www.devalspa.it/].

[38] The electricity consumption in each province published by Terna, [Online]. Available: [https://www.terna.it/it/sistema-elettrico/statistiche/pubblicazioni-statistiche].

[39] Concentration level of final users released by Italian Regulator [[19] Arera; Delibera 30 novembre 2017 793/2017/R/eel. "Determinazione dei premi e delle penalità relativi alla regolazione output-based del servizio di distribuzione dell'energia elettrica, per l'anno 2016"; https://www.arera.it/it/docs/17/793-17.htm].

[40] Textbooks and literature reporting statistics on Italian distribution systems P. C. N. Faletti; "Trasmissione e distribuzione dell'energia elettrica Vol. I"; Bologna: PATRON; 2004.

[41] F. Pilo et al., "ATLANTIDE — Digital archive of the Italian electric distribution reference networks," CIRED 2012 Workshop: Integration of Renewables into the Distribution Grid, Lisbon, 2012, pp. 1-4.

[42] E. S. Brancelli, «Il Sistema di Distribuzione di Enel Distribuzione,» Bari, 22 Marzo 2012.

[43] e-distribuzione; "Regole tecniche per le connessioni"; [Online]. Available: https://www.e-distribuzione.it/connessione-alla-rete/Regole_tecniche.html.

[44] ARERA, "Smart distribution system: promozione selettiva degli investimenti nei sistemi innovativi di distribuzione di energia elettrica", [Online]. Available: https://www.arera.it/it/eventi/150928.htm, 2015.

[45] G. Viganò, M. Rossi and D. Moneta, "Generation of artificial distribution networks for the evaluation of advanced control solutions," 2017 AEIT International Annual Conference, 2017, pp. 1-6.

[46] Number of primary substations per province extracted from [Ministero dello Sviluppo Economico (MiSE), "Piano Nazionale Integrato per l'Energia e il Clima", January 2020].

[47] G. Viganò, M. Rossi, C. Michelangeli and D. Moneta, "Creation of the Italian Distribution System Scenario by Using Synthetic Artificial Networks," 2020 AEIT International Annual Conference (AEIT), 2020, pp. 1-6.

[48] Правила о раду Дистрибутивног система, [Online]. Available: http://epsdistribucija.rs/pdf/Nacrt_PravilaORaduDS.pdf.

[49] Statistisk sentralbryå, [Online]. Available: https://www.ssb.no/.

[50] Tilastokeskus, [Online]. Available: https://pxnet2.stat.fi/.

[51] Statistics Sweden, [Online]. Available: https://www.scb.se/.

[52] Statistics Denmark, [Online]. Available: https://www.statbank.dk/.

[53] INEGI and APREN, "Energias Endogenas de Portugal," [Online]. Available: https://e2p.inegi.up.pt/index.asp], 2018.

[54] Bundesnetzagentur, List of Power Plants, [Online]. Available: https://www.bundesnetzagentur.de/DE/Sachgebiete/ElektrizitaetundGas/Unternehmen_Institutionen/Versorgungssicherheit/Erzeugungskapazitaeten/Kraftwerksliste/kraftwerksliste-node.html].

[55] Bundesnetzagentur, List of connected renewable energy sources, [Online]. Available: https://www.bundesnetzagentur.de/DE/Sachgebiete/ElektrizitaetundGas/Unternehmen_Institutionen/DatenaustauschundMonitoring/Marktstammdatenregister/MaStR_node.html.

[56] MFA, [Online]. Available: Ministero della Transizione Ecologica http://sinva.minambiente.it/mapviewer/index.html?collection=http://sinva.minambiente.it/WMC/Collection/VA/D5DB339E-DB14-9144-B043-B141DCABC108&v=full.

[57] ENTSO-E data, [Online]. Available:https://transparency.entsoe.eu/generation/r2/actualGenerationPerProductionType/show].

[58] Global Power Plant Database, [Online]. Available: https://datasets.wri.org/dataset/globalpowerplantdatabase].

[59] Italian Grid code, [Online]. Available: https://www.terna.it/en/electric-system/grid-codes/italian-grid-code].

[60] FlexPlan Consortium, 'Matched data', 'power plant list' [data sets], 2020.

[61] Energistyrelsen: "Data for eksisterende og afmeldte anlæg". [Online]. Available: https://ens.dk/service/statistik-data-noegletal-og-kort/data-oversigt-over-energisektoren (Accessed 23. May 2021).

[62] Sigrid Hendriks Moe (SSB), and Hanne Birgitte Laird (Miljødirektoratet): "Energiintensiv industri En beskrivelse og økonomisk analyse av energiintensiv industri i Norge", report no. 69/2013, Norges vassdrags- og energidirektorat, 2013.

[63] European Commission, "EU Reference Scenario 2016 Energy, transport and GHG emissions trends to 2050," Publication Office of the European Union, Luxembourg, 2016.

[64] IPCC, "2006 IPCC Guidelines for National Greenhouse Gas Inventories," IGES, 2006.

[65] Eurostat, "Energy balance June 2020 edition," 2020.

[66] European Commission, "Regulation (EC) No 166/2006 of the European Parliament and of the Council of 18 January 2006 concerning the establishment of a European Pollutant Release and Transfer Register and amending Council Directives 91/689/EEC and 96/61/EC," Official Journal of the European Union, vol. L 33, no. 4.2.2006, p. 1–17, 2006.

[67] European Commission, "Directive 2010/75/EU of the European Parliament and of the Council of 24 November 2010 on industrial emissions (integrated pollution prevention and control)," Official Journal of the European Union, vol. L 334, no. 17.12.2010, p. 17–119, 2010.

[68] European Commission, "Guide on EU corporate registration, third country and global registration under Regulation No 1221/2009 of the European Parliament and of the Council on the voluntary participation by organisations in a Community eco-management and audit scheme (EMAS)," Official Journal of the European Union, vol. L 330, no. 14.12.2011, p. 25–38, 2011.

[69] European Environment Agency, “EMEP/EEA air pollutant emission inventory guidebook 2019,” ISSN 1977-8449, 2019.

[70] Ramboll, CAMx User’s Guide Version 7.00, Novato, CA: Ramboll, 2020.

[71] A. Dunker, G. Yarwood, J. Ortmann and G. Wilson, “The decoupled direct method for sensitivity analysis in a three-dimensional air quality model - implementation, accuracy, and efficiency,” Environmental Science & Technology, vol. 36, no. 13, pp. 2965-2976, 2002.

[72] ECMWF, “ERA5,” [Online]. Available: https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5.

[73] ECMWF, IFS Documentation CY47R1, ECMWF, 2020.

[74] J. J. P. Kuenen, A. J. H. Visschedijk, M. Jozwicka and H. A. C. Denier van der Gon, “TNO-MACC_II emission inventory; a multi-year (2003–2009) consistent high-resolution European emission inventory for air quality modelling,” Atmospheric Chemistry and Physics, vol. 14, no. 20, p. 10963–10976, 2014.

[75] M. Guevara, C. Tena, M. Porquet, O. Jorba and C. Pérez García-Pando, “HERMESv3, a stand-alone multi-scale atmospheric emission modelling framework – Part 1: global and regional module,” Geoscientific Model Development, vol. 12, no. 5, p. 1885–1907, 2019.

[76] Eurostat, “Population on 1 January by age and sex, Eurostat - Data Explorer,” 2021. [Online]. Available: https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=demo_pjan&lang=en. [Accessed 28 06 2021].

[77] World Health Organization, “Health risks of air pollution in Europe - HRAPIE project - Recommendations for concentration–response functions for cost–benefit analysis of particulate matter, ozone and nitrogen dioxide,” [Online]. Available: www.euro.who.int/__data/assets/pdf_file/0006/238956/Health_risks_air_pollution_HRAPIE_project.pdf], 2013.

[78] Simbench Documentation, [Online]. Available: https://simbench.de/wp-content/uploads/2020/01/simbench_documentation_en_1.0.0.pdf].

[79] S. Meinecke, D. Sarajlić, S. R. Drauz, A. Klettke, L.-P. Lauven, C. Rehtanz, et al., “SimBench—A Benchmark Dataset of Electric Power Systems to Compare Innovative Solutions Based on Power Flow Analysis,” Energies, vol. 13, no. 12, p. 3290, Jun. 2020.

[80] C. Spieker, J. Teuwsen, V. Liebenau, S. C. Müller and C. Rehtanz, "European Electricity Market Simulation for Future Scenarios with High Renewable Energy Production", PowerTech 2015, June 2015.

[81] D. Menniti, M. Merlo, N. Scordino, N. Sorrentino, F. Zanellini, “A DSO-oriented mathematical model for dispersed generation management on MV networks,” 2012 IEEE Power and Energy Society General Meeting, pp.1,8, 22-26 July 2012.

[82] FlexPlan Consortium, “D1.1 Monte Carlo scenario generation and reduction” 2020.

[83] E-Redes “Rede Nacional de Distribuição – Situação a 31 de dezembro de 2020”. [Online] available at: https://e-redes-rede.wntech.com/

[84] "El sistema eléctrico español – Informe 2019", 2020, REE, [Online] available at: https://www.ree.es/es/datos/publicaciones/informe-anual-sistema/informe-del-sistema-electrico-espanol-2019].

[85] "Información básica de los sectores de la energía, 2007, Anexo", CNE, 2008.

[86] Iagua, [online: https://www.iagua.es, last accessed: 18/06/2021].

[87] "Las energías renovables en el sistema eléctrico español – informe 2019", REE, 2020, [Online] available at: https://www.ree.es/es/datos/publicaciones/informe-de-energias-renovables.

[88] M. Rylander and J. Smith, "Stochastic Approach for Distribution Planning with Distributed Energy Resources," CIGRE Grid of the Future Symposium, Kansas City, MO, Oct. 2012.

[89] EPRI, "Impact of High-Penetration PV on Distribution System Performance," Palo Alto, CA, 2011.

[90] Energistyrelsen: "Data for eksisterende og afmeldte anlæg", [Online] available at: https://ens.dk/service/statistik-data-noegletal-og-kort/data-oversigt-over-energisektoren, Accessed 23. May 2021.

