Synthetic Models of Distribution Networks Based on Open Data and Georeferenced Information

Giuditta Pisano [1],* , **Nayeem Chowdhury [1,2]** , **Massimiliano Coppo [3]** , **Nicola Natale [1]** , **Giacomo Petretto [2]** , **Gian Giuseppe Soma [1]** , **Roberto Turri [3]** and **Fabrizio Pilo [1]**

[1] Department of Electrical and Electronic Engineering, University of Cagliari, 09100 Cagliari, Italy;
    nayeem.chowdhury@enel.com (N.C.); nicola.natale@unica.it (N.N.); giangiuseppe.soma@unica.it (G.G.S.);
    pilo@diee.unica.it (F.P.)
[2] Enel Produzione Società per Azioni (S.p.A), 56122 Pisa, Italy; giacomo.petretto@enel.com
[3] Department of Industrial Engineering, University of Padova, 35131 Padova, Italy;
    massimiliano.coppo@unipd.it (M.C.); roberto.turri@unipd.it (R.T.)
* Correspondence: giuditta.pisano@unica.it; Tel.: +39-070-675-5876

Abstract: Many planning and operation studies that aim at fully assessing and optimizing the performance of the distribution grids, in response to the current trends, cannot ignore grid limitations. Modelling the distribution system, by including the electrical characteristics of the network (e.g., topology) and end user behaviors, has become complex, but essential, for all conventional and emerging actors/players of power systems (i.e., system and market operators, regulators, new market parties as service providers, aggregators, researchers, etc.). This paper deals with a methodology that, starting from publicly available open data on the energy consumption of a region or wider area, is capable to obtain reasonable load and generation profiles for the network supplied by each primary substation in the region/area. Furthermore, by combining these profiles with territorial and socio-economic information, the proposed methodology is able to model the network in terms of lines, conductors, loads and generators. The results of this procedure are the synthetic networks of the real distribution networks, that do not correspond exactly to the actual networks, but can characterize them in a realistic way. Such models can be used for all the kind of optimization studies that need to check the grid limitations. Results derived from Italian test cases are presented and discussed.

Keywords: modelling; synthetic networks; distribution networks; distributed generation; renewable energy sources; representative distribution networks; open data

1. Introduction

There is a general consensus on the fact that in the last 5–10 years, emerging trends have been radically changing the power system. One trend is the increasing connection of renewable energy sources (RES)—intrinsically intermittent and not programmable—to reducing CO_2 emissions and reaching climate goals. This trend, at the power distribution level, results in the increasing penetration of distributed generation (DG). Another trend at the distribution level is represented by new electric loads with high coincident peaks (e.g., electric vehicles, heat pumps, induction cookers, etc.) that impact the exploitation of existing distribution assets. In this context, the balance of generation and demand at every single point in time becomes more challenging and increases the final cost of electricity for the final users.

For fully assessing and optimizing the performance of distribution in response to the current trends, planning and operation studies that include the electrical characteristics of the network (e.g., topology) and end user behaviors have become complex, but essential, for all conventional

and emerging actors/players of power systems (i.e., system and market operators, regulators, new market parties as service providers, aggregators, researchers, etc.). Unfortunately, the optimization studies that aim at general validity results are affected by several practical issues, essentially due to the extreme variability of distribution systems, even at the regional/national scale. In fact, worldwide, the distribution systems are characterized by different voltage levels, different number of phases (single- or three-phase medium voltage MV distribution networks), different topologies constituted by many and many kilometers of lines of different kinds of conductors, different range of line lengths, supplied by a big number of primary substations that deliver energy to different categories of customers, whose data are often difficult to obtain for privacy protection. On the other hand, the studies that aspire to reproduce only specific real cases of given distribution networks are also seriously hampered by the limited observability of distribution systems and by the unavailability of data relevant to both customers and networks. For the majority of the power system players, different from distribution system operators (DSOs), as researchers, transmission system operators (TSOs), aggregators, etc., the data of distribution is completely unaffordable. Particularly, data on distribution systems are not open, even though known by the DSO, and very seldom the topology and the characteristics of a given network can be available for any studies outside the DSO departments.

Indeed, even if to some extent network topology may be represented for certain testing purposes, privacy regulations and market competition from the energy traders' side strongly stand in the way of the availability of consumption and production profiles at the distribution level and they are known only by energy suppliers and aggregators for the restricted group of their customers. This fact is further worsened by the significant amount of uncontrollable RES connected to the distribution systems that make it even more difficult the forecasting of the power exchange profile with the grid at the higher voltage level, by making the TSO/DSO interactions more difficult than in the past.

For the above-mentioned reasons, general guide and targeted methodologies for modelling distribution networks suitable for planning and operation purposes become the desiderata of all the players of power systems. The desired network model should be sufficiently general so that its application would not be hindered even to totally unknown networks, but also it should allow for capturing the specificities of a given distribution network of which some characteristics are known.

In the literature, for research purposes, benchmarking networks have been widely used. The most popular network models are the IEEE and CIGRE test networks, both for transmission and distribution systems. They are currently used by many researchers for different kinds of studies for testing optimization and control algorithms, reliability calculations, power quality issues, etc. [1,2]. The IEEE networks do not exactly match with the European distribution system, which is quite different from the US one, for (un)balanced loads and networks. European and US versions of CIGRE benchmark systems were developed for calculating network integration of RES and distributed energy resources (DER). Despite their undeniable contribution, IEEE and CIGRE test networks are benchmark networks that can be used for general purposes but may hardly be adapted to assume some specific characteristics of a real given network. Alternatively, for demonstrating the effectiveness of their proposed solutions, many authors use test MV and LV distribution networks derived from real cases or ad hoc designed for emphasizing critical conditions that can be solved with their control and optimization algorithms, able to find the optimal planning alternative for the distribution network development, or the optimal scheduling of the distributed energy resources for minimizing operation costs, or the optimal position of devices useful to a wide range of purposes (see, for instance, References [3–8]). Few research projects proposed a repository of representative networks of a given national ambit (for instance in Italy [9] and in the UK [10]). Recently the European Commission published the Distribution System Operators Observatory 2018 [11], which aims at giving an overview of the electricity distribution system in the EU. From the report, arisen from a survey carried out on many European DSOs, emerges the diversity of the characteristics of the existing distribution networks in Europe. Within this project, large-scale and feeder-type network models, based on real technical data provided by the DSOs, have been proposed as the first step for obtaining European representative networks, including large-scale distribution

systems [12,13], which proposes a data-driven topology estimation method for medium and low voltage (MV and LV) urban distribution grids by utilizing the historical smart meter measurements and by capturing statistical dependencies amongst bus voltages using a probabilistic graphical model. In that paper, the position of the nodes is known, and the voltage profiles are available. The use of spatial databases for assigning geographical coordinates of the distribution network nodes and the resort to geographical information systems have been successfully proposed in several papers, for building a model of unknown existing networks or, even, for helping planners to choose a suitable route for new lines [12,14].

This paper deals with a methodology that, starting from publicly available open data on the energy consumption of a region or wider area, is capable to obtain reasonable load and generation profiles for the network supplied by each primary substation in the region/area. Furthermore, by combining these profiles with territorial and socio-economic information, the proposed methodology is able to realistically model the grid in terms of lines, conductors, loads, and generators. The network model is developed by resorting to typical feeders, representative of given ambits, and it allows for calculating possible outbound conditions (e.g., excessive voltage variations, excessive power flows, etc.), that, finally, can be fixed by exploiting smart control systems able to find the optimal operation of networks.

The main contribution of the paper is the methodology for modelling real distribution systems, both in terms of demand and production profiles and in terms of lines and conductors, by using open data only and representative network portions. The proposed methodology may be useful to evaluate, for a given portion of the distribution grid that covers quite a wide area, the impact of new regulations or market rules, to validate the effectiveness of smart control and management systems, to estimate the opportunity to participate in the markets, etc.

In this paper, the methodology is applied to the Italian case, but it is applicable to any other world case simply by slightly changing the representative network portions to make them suitable to other national contexts.

The structure of the paper is as follows: Section 2 describes the characteristics of the distribution system model that can be obtained as final result of the methodology. Section 3 deals with the open data useful for the purpose. Section 4 describes the representative network portions adopted for the grid modelling. Section 5 describes the Italian test case, and finally, Section 6 is about remarks and conclusions.

2. Distribution System Model

In the paper, for the sake of simplicity and without lack of generality, the real distribution systems are supposed as interfaced with the high voltage (HV) grid by the primary substation (PS), that represents the point of common coupling where the MV distribution system (DSO managed) is connected to the HV transmission system. Such a TSO/DSO interface is constituted by one or more dedicated HV/MV transformers. The position of the TSO/DSO interface has to be adapted to a specific grid code or regulatory framework but, in the case of the paper, the transformers are owned by the DSO, thus, the TSO/DSO interface is the HV node. In other cases, the transformer may be owned by the TSO, so the TSO/DSO interface becomes the MV busbar of the transformer. Whatever the TSO/DSO interface position, before the massive connection of generation to distribution systems, the energy flowed uniquely from transmission to distribution throughout that interface. Nowadays, at the interface point, bidirectional flows of energy are frequent, and energy can often flow from distribution networks towards the transmission system. A good equivalent model that represents the distribution system behind the TSO/DSO interface is a power plant that could theoretically operate on four quadrants by including both positive and negative values of active and reactive powers, with a circular capability curve or even rectangular, if active power P and reactive power Q can simultaneously assume their maximum values (Figure 1). In practice, the reactive power flow through the interface is often only inductive ($Q > 0$) and thus, during the year, the setpoints of the equivalent power plant are all placed

in the right two quadrants, that represent positive reactive powers. For instance, in Figure 2, each dot represents a couple of (P, Q) that characterizes the TSO/DSO interface in a generic time interval of the year (i.e., 1 h) of a given PS that supplies a network on which many DG plants, renewables or not, have been connected. The colors indicate the typical day relevant to the time interval. The dots in the negative half-plane of P signify a reverse flow towards the HV system. The capability curve of such a power plant can be considered the envelope of the area covered by these points.
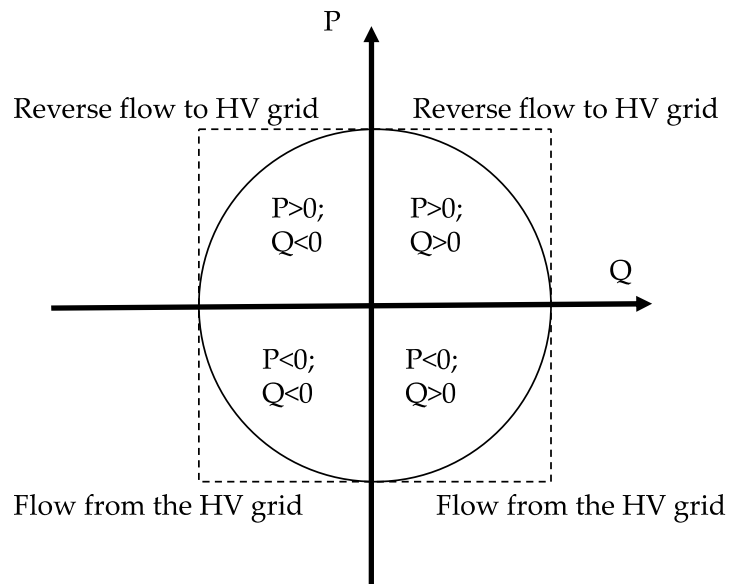

Figure 1. Theoretical four quadrants capability curve of the virtual power plant representing the primary substation.
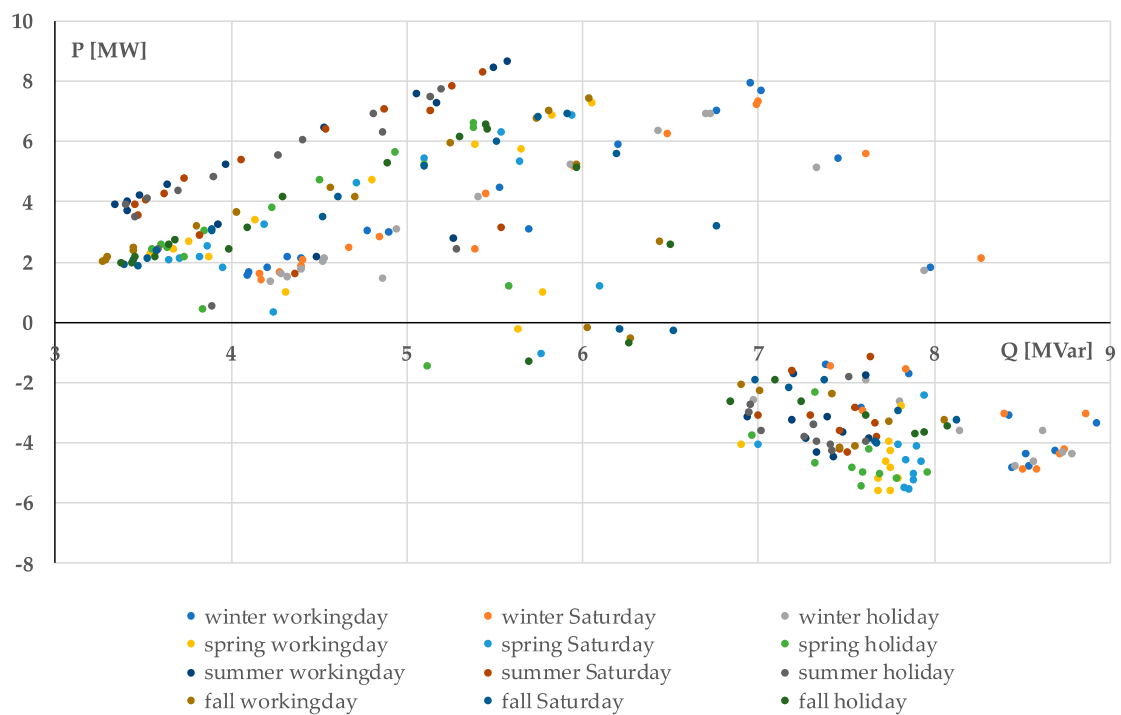

Figure 2. Setpoints at the transmission system operator (TSO)/distribution system operator (DSO) interface during typical days in the year.

The prediction of the production/consumption from/to such a particular virtual power plant is limited by the limited observability of distribution systems worsened by the intrinsically low programmability of the significant amount of RES downstream of the TSO/DSO interface. Thus, to fill this gap in observability, the active and reactive power profiles at the interface have to be estimated, trying to eliminate the uncertainties and to make the estimation as realistic as possible.

The estimation of the active and reactive profiles at the TSO/DSO interface could be enough to improve the observability of the distribution systems by making less uncertain the power exchange through the transformers at the TSO/DSO interface. No changes in the expected profiles mean to bring back the power system management to the era before massive RES and DG diffusion when the TSOs were able to foresee the demand of the distribution systems with a very low level of uncertainty. Indeed, a smaller error in the forecast of the residual load demand can reduce the need for balancing services and, consequently, the ancillary service costs, related to the significant role of RES connected to the distribution.

In the proposed methodology, such estimation has been made with considerations and calculations that arise from global data of electric demand and production of a given region or area, provided by public or semi-public bodies involved in the transparency activities related to the electricity and from other assumptions regarding territorial and socio-economic information. This information is related to the demography, the land cover and usage, as, for instance, the number of buildings, the number of employees, etc., and can be useful to characterize the area served by a given PS, as detailed in the following Section 3.

Nevertheless, the exchange profiles are not the only figures needed to model distribution systems. The network model in terms of topology, lines, and position of loads and generators is important to foresee any possible critical condition that may alter the regular operation of the distribution systems. Such conditions may arise, for instance, from changes in the expected setpoints of DERs resulting from the participation of DERs in the ancillary service (AS) market. The intent, manifest in the recent proposals of the regulatory bodies, of exploiting the flexibility potentially offered by the distribution networks in the AS market to contribute in reducing the costs of the services, should imply that distribution networks will be able to tolerate small or big variations of the expected setpoints of the connected DERs that want to offer such services to the AS market [15–17]. Unfortunately, despite the fact that the DERs have been connected under the hypertrophic conditions generated by the fit and forget approach, distribution systems are no longer robust enough to accept any changes in DER production/consumption without operational and control actions. The estimation of the distribution flexibility cannot disregard the grid technical limits, and detailed grid models of the distribution system are required. However, for the majority of the power system players, different from DSOs (e.g., researchers, TSOs, aggregators, etc.), the data on the topology and the characteristics of a given network distribution are not open and completely unaffordable for any studies outside the DSO departments. Whereas such data would be very useful for TSOs and other players, as aggregators of private DERs, that need a priori knowing the impact on the distribution system operation of the changes derived by the bids that might be awarded in the market. On one hand, the TSOs, that should take advantage from the flexibility potentially offered by demand and DG to reduce quantities in the AS market (e.g., for solving balancing problems, increasing reserve capacity, and other services), have to avoid any erroneous evaluations on the availability of flexibility services from distribution, and, on the other hand, the aggregators have to be confident about the market potential of the customers that belong to their portfolio to prevent any possible technical limitations in the feasibility of their offers. In addition, also the DSOs, even they know their own networks, would like to perform studies on models of their networks able to give results which are easily generalizable for massive applications, aimed at evaluating the potential of the DERs to sustain the operation of the local system (via distribution networks active management) and to contribute in solving some problems related to the interaction with the higher voltage level grid, without undermining the secure operation of the local systems.

For all these reasons, modelling not only the exchange profiles at the TSO/DSO interface, but also the distribution systems, in terms of topology, lines, conductors, loads, and generators, for foreseeing the operation problems that could arise by changing DER setpoints, becomes fundamental for many activities relevant to the new era of flexibility. The proposed methodology for modelling the distribution system by using representative network portions is detailed in Section 4.

3. Modelling Energy Consumption and Production at the Distribution Level with Open Data

In this work, the greatest attention has been paid to develop a consistent and repeatable methodology that uses only publicly available open data since not all players have access to classified information. In fact, from the energy traders' side, privacy regulations and market competition strongly limit the availability of consumption and production profiles at the distribution level and they are known only by energy suppliers and aggregators for the restricted group of their customers. However, some data about the transmission power systems and about the territory and the socio-economic situation are open or saved in repositories and databases that can be freely requested, or in a few cases, purchased from semi-public institutions. Obviously, these data are not ready to be used as they are, but if suitably managed and combined, they may be a valid help for modelling the behavior of distribution networks. The proposed methodology exploits even raw open data, processed for extracting the maximum level of information, but, clearly, the more precise and ready the input the more reliable the results Furthermore, the procedure is also able to reach the goal if it is available a limited quantity of data, but, in this case, it is necessary hypothesizing the most important missing data. In practice, for obtaining reliable and accurate results, the total demand and production of the whole area, the PS locations, the share of population in the portions on which is subdivided the area, and, at least, the rough land cover for assigning the ambit (rural, industrial and urban) are needed. On the contrary, if in a given case there is also specific information available, generally unknown, the procedure can properly include it (e.g., the consumption/production profile of one load/generator known by a measurement campaign).

In the following, explicit reference to open data available in Italy will be made, but similar data are available for EU and non-EU countries.

Generally, information about energy consumption and production is available at regional or, even bigger, territory level. Thus, the first goal is the assignment of the right sharing of consumption and production of a given region to each PS included in the area, starting from this global information. This activity is necessary due to the above-mentioned limited observability of the distribution system, particularly with reference to generation data and for the restrictions on the use of the database with customer's behavior for privacy policies. In this paper, aggregated data on energy consumption and production in a given region or market zone have been used, since they are available on a yearly basis and published by Italian TSO and authorities, as the national electricity market operator [18,19]. In addition, some information on customers and producers directly connected to the transmission network are known for the Italian National Transmission Grid (NTG) and they are regularly published in the Italian TSO website [18].

Figure 3 summarizes the methodology for the assignment of generation/consumption to each PS that is described below.
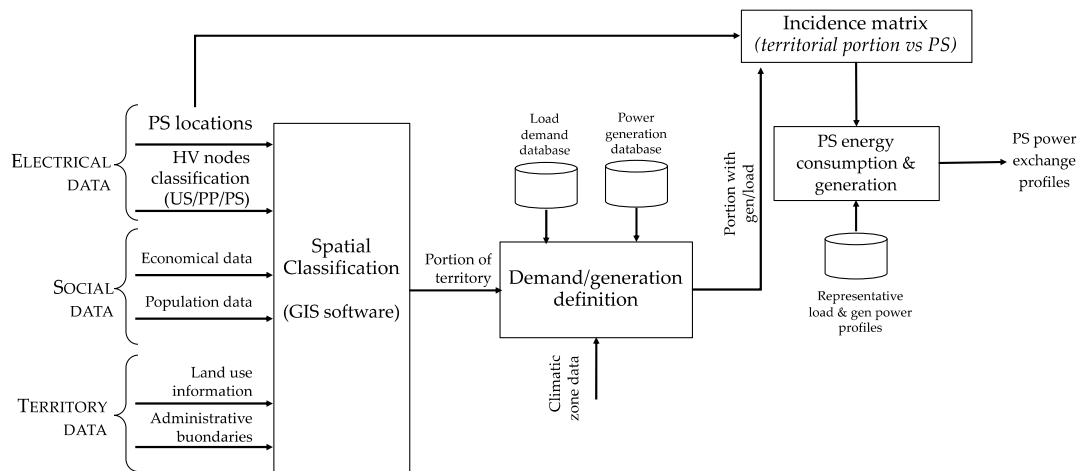
Figure 3. Procedure for assigning load/generation profiles to each primary substation (PS) in a given region (flow diagram).

The procedure, that is also pictorially depicted in Figure 4, can be split into the following tasks:

1. Build databases containing the PS locations of the region/area, and subdivision of the NTG HV nodes into HV customers (user substation—US), pure producers (PP) and public distribution substations (PS themselves).

2. Spatial classification: subdivide the region/zone into portions, gradually smaller, according, for instance, to the administrative boundaries, and combine with such subdivision the information regarding the territory and the population by resorting geographical information systems (GIS) applications and tools.

3. Assign to each territorial portion the share of demand consumption and production of the region/area.

4. Build an incidence matrix associating the territorial portions to one or more PSs.

5. Assess the total energy consumption and generation for each PS.

6. Define the active power time series for each PS.

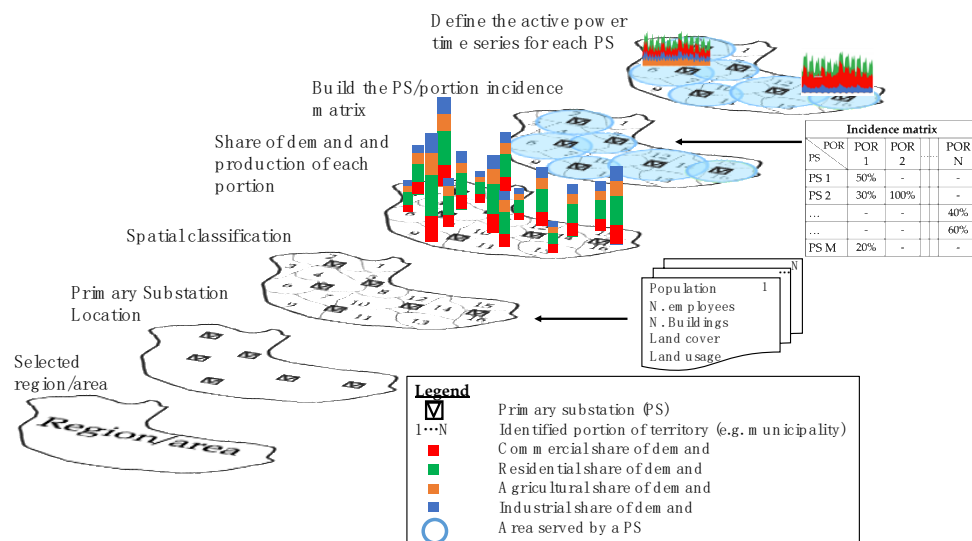7. Associate the geographical information with each PS, for the grid modelling (see Section 4).

Figure 4. Illustration of the steps of the procedure for assigning load/generation profiles to each PS in a given region/area.

In item 1, a list of PS of the given region is drawn up. The location of each PS is generally unknown but, in some cases, there are maps and shapefiles that can be downloaded from GIS applications published by local public bodies (i.e., the Italian National Geoportal that is an access point to environmental and territorial information [20]). Furthermore, the identification of the public, private, or another role of each PS has been defined. In fact, different kinds of PS have been addressed in different ways. The main attention in this paper is paid to model the public PS, that supply distribution networks, but for contributing to the next tasks of the items 3 and 5, the energy demand of the user substations has been preventively estimated by assuming a specific consumption of energy per production unit, diversified for the industrial sector. Table 1 reports some specific energy consumptions per production unit assumed in this paper for the most common industrial sectors. More in detail, for each US that falls in a specific area, firstly, the estimation about its yearly production or its time of use is needed. Then, by considering this information and by multiplying it by the data in Table 1, the energy demand of a given US can be assessed. Once estimated, the consumption of a given US is deducted from the energy delivered of the territorial portion which it belongs to, as described below (the items 3 and 5).

Table 1. Specific energy consumptions per production unit.

Industrial Sector	Specific Energy Consumption
Water	120 (kWh/t)
Cement	67.5 (kWh/t)
Chemical	67.5 (kWh/t)
Refinery	67.5 (kWh/t)
Metallurgy	500 (kWh/t)
Paper	900 (kWh/t)
Mining	300 (kWh/t)
Mechanical	10 (MWh/h)
Electronic	10 (MWh/h)

Regarding item 2, named spatial classification, the territory is initially subdivided into portions gradually smaller, according, for instance, to the administrative borders. In Italy, the regions are constituted by provinces and the provinces by municipalities. The dataset with information relevant for the area under investigation is generally formed by merging open data sources that are prepared by national, regional, provincial, and municipal institutions. The granularity of geo-spatial analysis is the municipality dimension. Open data are not all ready to be used in GIS software for geographic analysis and a pre-processing activity is necessary to adapt data format and to upload relevant information in the GIS software and to populate and build GIS maps with different layers of georeferenced data (e.g., demographic, economic, and social variable maps) for the territorial portions on which the region has been discretized. With particular reference to the example presented in the paper, the GIS maps published by public institutions with layers on the land cover and usage and the distribution of commercial and industrial buildings have been used [20]. In more detail, the map of each territorial portion has been enriched with layers on demographic density, on the typology and the destination use of buildings (e.g., as industrial or commercial), the number of employees per commercial/industrial sector, and the land cover categories. It is worth noticing that the land cover categories do not exactly correspond to the three ambits (i.e., industrial, urban, and rural) that will be used for classifying the network (item 7 and Section 4). For this reason, a clustering procedure has been done to produce classes that can be associated to one of three ambits (as detailed below). For such purpose, the dataset derived from the Corine Land Cover (CLC), that provides from 1985 consistent localized geographical information on the land cover of some member states of the European Community, may also be

useful [21,22]. At the end of this step, a high spatial resolution classification of each elementary area is provided. For instance, regarding the population, the share of the population of each municipality in percentage of the total population of the region assigned in this step to the territorial portions is used in item 3 to attribute the residential consumption of electricity, as detailed below.

The information regarding the population and, in general, the territory, have been added to this study in order to tune the share of energy associated to each territorial portion (item 3), by starting from the total energy demand and generation of the region and according to demographic, economic, and social variables typical of the given territorial portion included in the region. In more detail, item 3 exploits the administrative subdivisions of the regional territory. For the scope, demography, and employment, information has been gathered for any municipality from the Italian National Statistics Institute (ISTAT) [23]. The goal is to attribute the right, or the most realistic, share of the total demand of energy to each municipality or smaller portion of territory. This can be made by using a top-down approach that, starting from the gross data on energy consumption and production in a given region, considers the specificities of the territory as they result from the maps of item 2 and from the demographic and economic information derived from Reference [23] and the other used databases [20].

The data on the energy consumption of a given region or market zone are generally subdivided into four main end-user categories: commercial, industrial, agricultural, and residential, used in this paper to identify the energy share assigned to each municipality. In particular, the number of commercial and industrial buildings, the number of the agricultural holdings, and the data of the employees of a given area, differentiated by product or service category, have been used for allocating the industrial, agricultural, and commercial electric demand, respectively. Instead, the population may be properly used to assign the residential consumption to each municipality. The percentages of these quantities assigned to each municipality on respect to the total referred to the whole area/region are used for identifying the share of consumption of a given territorial portion.

Furthermore, it is worth noticing that the presence in the municipality territory of a US, identified by item 1, increases the share of the industrial demand of that municipality, compared to the others of the region. In addition, if in a given area, PSs that supply railway transportation have been identified, data about the energy consumption at the regional or provincial level, made available by the Italian TSO [18], have been distributed among the railway PS, adopting the population that may use the railway service as weights. In such a way, the railway substations that are close to the most crowded areas (as cities or the urban areas) have been estimated to consume more energy than the ones that are located in sparsely inhabited zones.

Regarding the generation connected to the distribution grid, since the vast majority of DG is based on RES or high-efficiency CHP, open data exist on the position and size of those generators, due to public incentive programs [24]. Thus, the production by the local generation has been associated with each municipality according to such information, by considering the rated power capacity of the generators and the production potential of the site.

At the end of item 3, each territorial portion has been characterized by the share of demand consumption and production of the area. The need, at this point in the procedure, consists of creating an incidence matrix that associates each municipality area to one or more PSs. This task is solved by using a GIS software procedure, ad hoc implemented (item 4). The creation of the incidence matrix is crucial for assigning the proper generation/consumption profile to each PS. This matrix allows splitting the regional (zonal) global consumption/generation to each PS (item 5). This assignment is not straightforward because the loads actually served by each PS is unknown, and it is difficult to univocally attribute some parts of the territory to one or to another PS, especially if they are positioned in a zone equidistant from two or more PSs. Thus, to make realistic assumptions, the proposed procedure assigns, first, a circular area around the PS, by considering a first attempt radii r (e.g., the smallest distance d_{min} between the PS and the adjacent ones). In a second step, the circles have been

adjusted according to the following heuristic rules, formulated by considering the typical lengths of the distribution networks lines:

- The radii of the circles around PSs predominantly urban have been increased, due to the too-small distance between them (e.g., if $d_{min} < 1$ km, $r = 3 \cdot d_{min}$);
- On the contrary, the radii of the circles around the PSs mostly rural have been reduced (e.g., if $d_{min} > 16$ km, $r = d_{min}/2$).
- The radii of the circles around the PS different from public (i.e., USs, railway substations, etc.) have been chosen equal to 250 m.

Then, the center positions and the radius lengths have been further corrected with an iterative process that converges when at least a small portion of each municipality is covered by the circles (see for instance a small territorial area of the Sardinia island covered by circles in Figure 5). The final segmentation of the territory around the PS may not be centered in the PS and not have a circular shape. In case more than one PS has been associated with the same municipality territory, the GIS software also helps to find the influence area of each PS by calculating the sizes of the intersected areas.
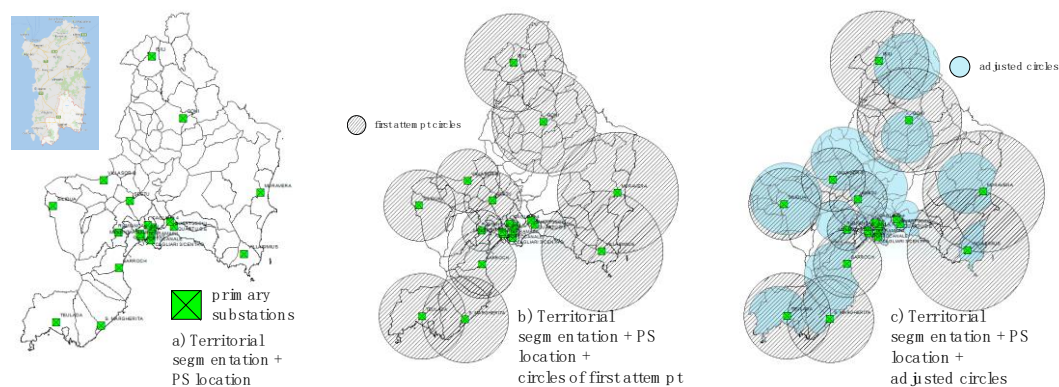

Figure 5. Procedure for assigning for creating an incidence matrix to associate each municipality area with one or more PSs: (**a**) Territorial segmentation + PS location, (**b**) circles of the first attempt, and (**c**) adjusted circles.

At the end of this process, the global yearly energy consumption for agricultural, industrial, residential, and tertiary usages can be assigned to each PS in the region (item 5). Also, by correlating the data on distributed generation with the territory with GIS through the incidence matrix, yearly DG production and typology of generators can be assigned to each PS.

Anyway, global figures, as the yearly energy consumption and production, are not enough for modelling the behavior of the distribution system. Time series of consumption and production are necessary to properly represent the impact of demand coincidence and generation/load homotheticity (item 6). Thus, in this paper, hourly curves are built by using twelve daily power representative profiles that model the consumption in typical days of the year (i.e., working days, Saturdays, and holidays for the four seasons, respectively) for agricultural, industrial, residential, and tertiary (i.e., commercial and offices) loads, derived from a digital archive of Italian representative distribution networks [9]. Figures 6–8 show the typical daily profiles used for the agricultural, industrial, and tertiary customers, respectively. Such curves define the relative coefficients that, hour by hour, have to be multiplied by the annual peaks for obtaining the absolute consumption profiles. By observing the typical daily profiles, it is evident that while the peak of agricultural consumption occurs in the winter working day (i.e., when the curve reaches 1.0 p.u), it occurs in the summer working day for the other categories of customers (i.e., industrial and tertiary).
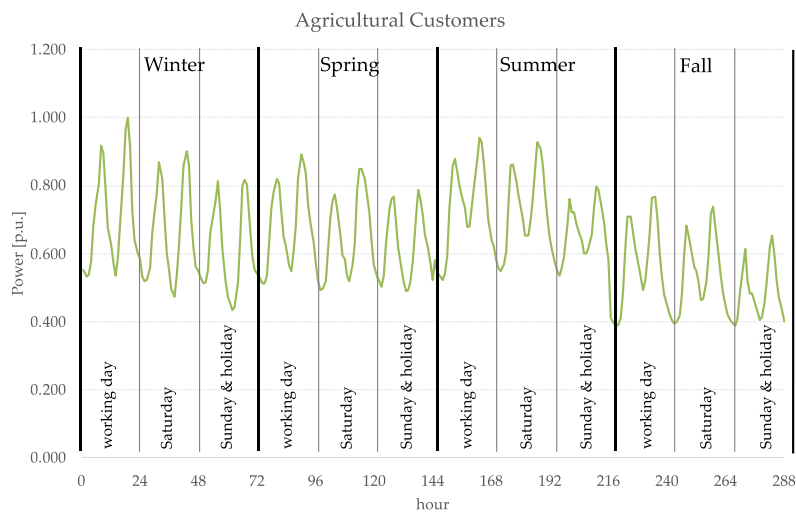
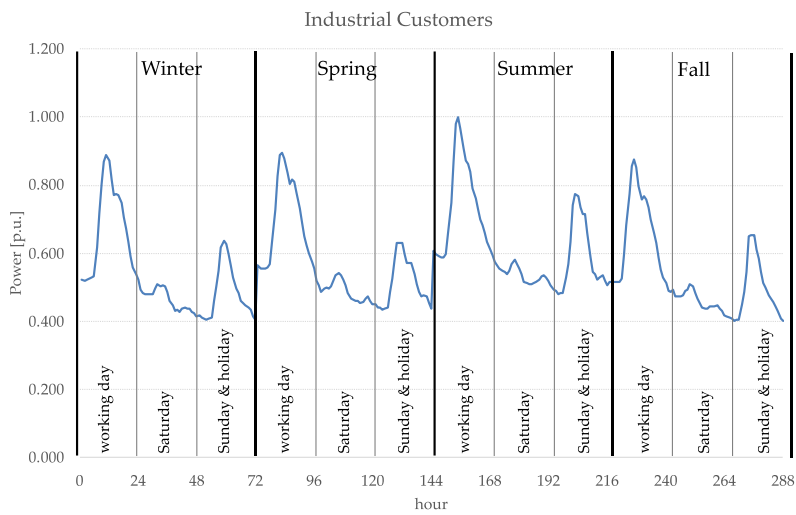**Figure 6.** Typical daily profiles used for the agricultural customers.

Figure 7. Typical daily profiles used for the industrial customers.
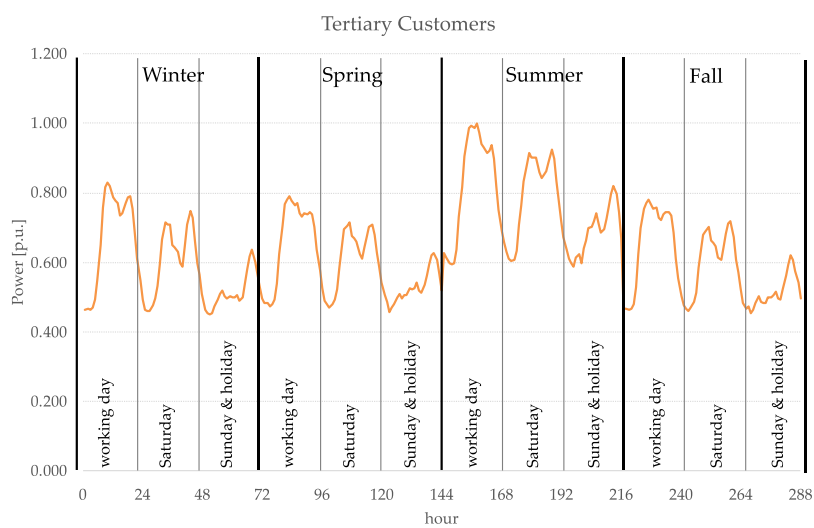

Figure 8. Typical daily profiles used for the tertiary customers.

Finally, a further specification is made for properly considering the difference of the electrical behavior of customers of different climatic zones and the difference of those production profiles that depend on the position of the considered region. In Italy, the "climatic zone" concept refers to a specific classification of the Italian national territory, identified by one of the first six capital letters (A–F), that attributes typical climatic characteristics to each Italian municipality independent from their geographical location, (The classification was made in order to limit the energy consumption by heating plants and depends on the degree-day, that is the sum, extended to all days in a conventional annual heating period, of positive differences between internal temperature (conventionally fixed at 20 °C) and the average daily external temperature. Depending on the climatic zone to which a municipality belongs, the period of the year and the maximum number of hours per day when heating may be switched on has been defined (e.g., municipalities belonging to the C zone have a degree-day between 900–1400 and may switch on heating for 10 hours a day in the period 15 November–31 March))(such subdivision is shown in Figure 9) [25]. In Figure 10, the typical daily profiles of residential customers differentiated by climatic zone are reported. The differences between climatic zones are more evident in summer when the electrical energy consumption of residential customers is strongly affected by the use of air-conditioning.
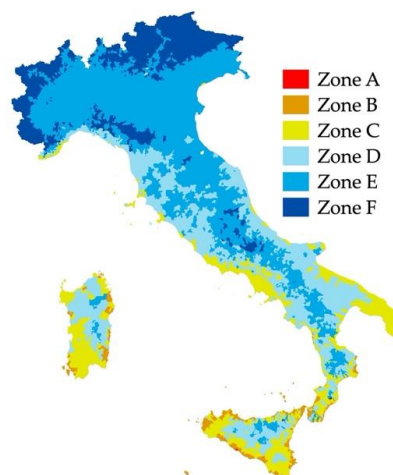

Figure 9. Subdivision of the Italian territory in climatic zones.
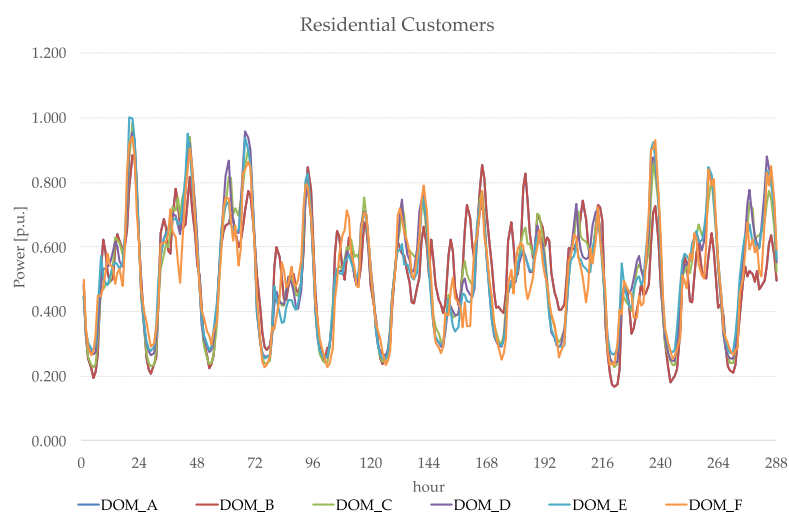

Figure 10. Typical daily profiles of residential customers differentiated by climatic zone.

The same approach is also used for the generation, by considering typical production profiles of photovoltaics (PV), wind turbines (WIND), hydro plants, and CHPs (combined heat and power). Such typical production profiles are derived by combining the technical skills of the specific power plants with the historical meteorological data for the considered region/area. In more detail, in the proposed methodology, the PV production, since the solar irradiation strongly depends on the latitude, has been distinguished in three latitude ranges that correspond to the North, Central, and South Italian territory, respectively. The wind production profiles have been differentiated between coastal and mountain installations, the CHP production profiles have been related, as the demand, to the climatic zone subdivision, and finally, a unique hydro-typical production profile has been assumed for the whole national territory.

In conclusion, referring to the entire methodology shown in Figures 3 and 4, heterogeneous data inputs the methodology at different stages for identifying the share of demand and generation amongst the PSs in the examined area. These shares, combined with representative time series for each category of customers and generators, allow for constructing the time-series of both production and consumption at each TSO/DSO interface, namely primary substation. The final task of the procedure (item 7) is made for identifying the characteristics of the network that supplies the territory served by each PS. In this task, the geographical information (e.g., land cover and usage, number of buildings, etc.) that has been associated with each territorial portion in item 2 and then assigned to each PS in item 5, has been elaborated for the network modelling described in the next Section.

4. Modelling the Distribution System with Representative Network Portions

As argued in Section 2, the estimation of the distribution network flexibility cannot ignore the grid limitations. Thus, a model that represents each network with lines, conductors, loads, etc., is fundamental. In transmission grids, the grid topology is usually available to system operators. The extreme variability of the distribution system has suggested the resort to representative networks with common features of distribution feeders useful for planning and operation studies. Papers and research projects proposed different models and techniques for producing representative and reference networks [12,26,27]. As in Reference [12], the final representative network is the result of a fine-tuning process, that can start from specific information from DSOs opportunely managed for extrapolating indicators useful for classifying networks with the same global characteristics (i.e., number of served customers, circuit length, covered area, etc.). This approach may exploit clustering techniques for identifying the most typical configurations. Otherwise, the fine-tuning process is used for planning activities and leads to networks that have the aim of connecting all the end-users by gathering their geographical information.

In this paper, representative networks derived from a clustering process that identified the most typical configurations are used for modelling real distribution networks starting from their territorial characteristics, acquired by GIS applications. No specific information about a single customer is needed, even if some data are known, they can be used for improving the model.

The authors, within the ATLANTIDE project (Archivio TeLemAtico per il riferimento Nazionale di reTI di Distribuzione Elettrica" that means "Digital archive for the national electrical distribution reference networks"), have co-authored a methodology and contributed to building a repository of representative networks for the Italian distribution system. The network topology, the feeder type (e.g., overhead and underground), the distribution of different customer categories (i.e., residential, industrial, agricultural, tertiary, transportation), and the daily consumption and generation profiles, etc., were used to build representative networks for urban, industrial, and rural areas [9]. Without a lack of generality, portions of the ATLANTIDE representative networks have been considered in the model proposed in this paper, but any other representative networks can be used instead of them (for instance, the models in References [27,28]). In particular, elementary portions of the ATLANTIDE representative networks have been used to build synthetic networks downstream of each PS that, as from the meaning of such kind of network models, do not correspond exactly to the actual networks

but can characterize them in a realistic way. Figure 11 depicts the developed approach within the GIS application: layers that include information about buildings, unelectrified areas (i.e., lakes, ponds, forests, etc.), land usage, and cover are superimposed on to each other for building synthetic networks that result from a combination of elementary network portions located in an integrated final layer. Such an approach can be summarized as follows:

- Association of the geographical/socio-economical information to the territory supplied by a given PS, by exploiting GIS tools and applications (shapefiles and intersections).
- Assessment of the shares of rural, industrial, urban ambit according to the land usage, population, etc., related to the territorial portion.
- Combining elementary portions of representative networks for building the synthetic network that models the grid downstream of the HV/MV transformer.
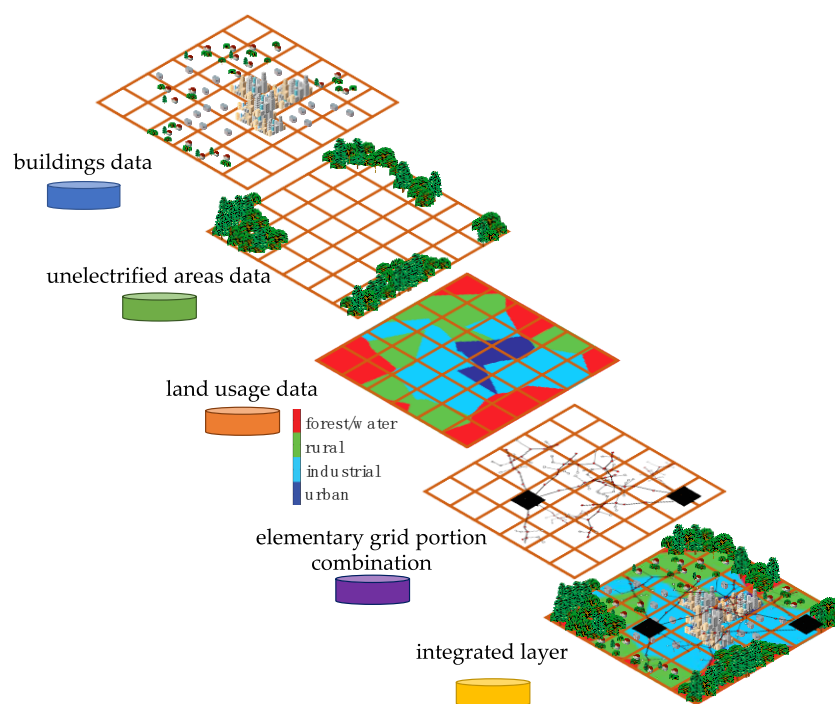

Figure 11. The procedure for building synthetic networks.

It is worth noticing that, as mentioned above, the land cover categories do not exactly correspond to the three ambits (i.e., industrial, urban, and rural), but it is possible to cluster such categories in classes that can be associated to one of these three ambits. For instance, land cover classifications like mineral extraction or dumpsites are identified as belonging to the industrial ambit, vineyards and olive groves to the rural ambit, and sport and leisure lands or shopping centers to the urban ambit.

The elementary network portion used to describe a distribution network supplied by a PS in a given area, whose consumption/generation patterns have been developed with the methodology in Section 3, is the representative feeder. Particularly, rural, industrial, and urban representative feeders have been defined under the following assumptions:

- Rural feeder: long overhead lines with lateral branches, low demand, and spread LV customers (low power density), mostly with agricultural or residential profiles.
- Urban feeder: relatively short underground cable lines, high power density, mainly LV residential loads and low total demand per feeder (but in an urban area, normally, a PS supplies quite a high number of feeders).

- Industrial feeder: supplies various different load types, high rate of MV loads, relatively short line extension, and quite high total power demand.

The choice of these three feeders among the others of the representative networks defined within the ATLANTIDE project has been driven by representativeness criteria (e.g., the maximum percentage of overhead lines for the typical rural feeder) and by other reasons related to requirements useful for suitably modelling any real PS with the same elementary portions and making the representation more flexible and scalable (e.g., energy consumption not too high for the typical industrial feeder). Furthermore, for representing the case of a dedicated connection for quite big power plants, another typical feeder is added to the previous list. Such a feeder does not supply any loads and it is supposed to be very short to avoid network losses increasing.

Tables 2 and 3 report the main characteristics of the representative feeders and their share of demand consumption. In Figure 12 the layout of the first three typical feeders is shown.

Table 2. Main characteristics of the representative feeders.

Feeder	MV Nodes	Total Length (km)	Max Distance from PS (km)	Load (MVA)	LV Installed Power (%)
Rural	22	40.11	20.85	3.51	99.6
Urban	9	1.23	1.21	3.62	97.6
Industrial	22	18.09	11.21	4.04	32.1
Dedicated	1	0.01	0.01	–	–

Table 3. Share of demand consumption (cons.) of the typical feeders.

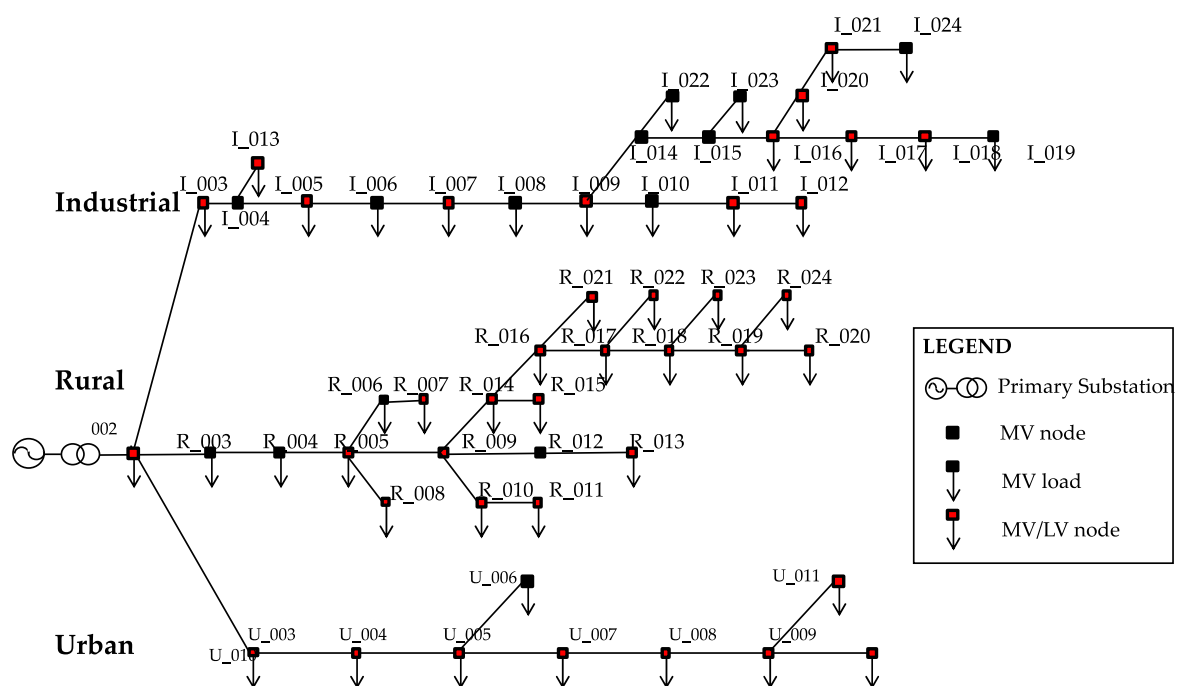| Feeder	Agricultural cons. (%)	Industrial cons. (%)	Tertiary cons. (%)	Residential cons. (%)	Total Energy Consumption (GWh/y)
Rural	37.16%	0.38%	0.00%	62.46%	16.98
Urban	0.00%	2.49%	55.78%	41.73%	13.54
Industrial	0.00%	65.91%	10.47%	23.62%	21.48

Figure 12. Layout of the three considered typical feeders.

Each PS model of the examined area is obtained with a combination of the four typical feeders previously described. The number of typical feeders composing the distribution network is determined by using the information gathered with GIS software from open databases that are of common use in civil and transportation engineering according to the results of item 7 of the procedure described in Section 3. As in Reference [26], the GIS information about land usage is used to classify the territory as urban, rural and industrial. The share amongst these ambits reflects the composition of the power system that supplies the region (e.g., in urban areas the use of buried cables is prevalent) and it is used to find the number of representative feeders of each category whose total energy demand minimizes the difference with the effective total demand assigned to the PS by the procedure described in Section 3 (item 5–6). The minimization is performed by solving the system (1–5) that gives the number of typical rural, urban, and industrial feeders composing the network model of a specific PS.

$$nF(\text{U}) \cdot LF(\text{U}) + nF(\text{R}) \cdot LF(\text{R}) + nF(\text{I}) \cdot LF(\text{I}) = L_{tot}(\text{PS}_k) \tag{1}$$

$$nF(\text{U}) \cdot LF(\text{U}) = \text{U}_\%(\text{PS}_k) \cdot L_{tot}(\text{PS}_k) \tag{2}$$

$$nF(\text{I}) \cdot LF(\text{I}) = I_\%(\text{PS}_k) \cdot L_{tot}(\text{PS}_k) \tag{3}$$

$$nF(\text{R}) \cdot LF(\text{R}) = R_\%(\text{PS}_k) \cdot L_{tot}(\text{PS}_k) \tag{4}$$

$$nF(\text{U}) \cdot EF(\text{U}) + nF(\text{R}) \cdot EF(\text{R}) + nF(\text{I}) \cdot EF(\text{I}) = E_{tot}(\text{PS}_k) \tag{5}$$

where:

- $nF(\text{U})$, $nF(\text{R})$, and $nF(\text{I})$ are the number of urban, rural, and industrial feeders respectively, each associated with the lengths $LF(\text{U})$, $LF(\text{R})$, and $LF(\text{I})$.
- $U_\%(\text{PS}_k)$, $R_\%(\text{PS}_k)$, and $I_\%(\text{PS}_k)$ are respectively, the urban, rural, and industrial shares of land cover and usage in the area supplied by the k-th PS.
- $EF(\text{U})$, $EF(\text{R})$, and $EF(\text{I})$ are the annual energy consumptions for each of the three typical feeders.
- $E_{tot}(\text{PS}_k)$ is the total annual consumption of the k-th PS as derived by the energy profile assessment defined in Section 3.

Finally, $L_{tot}(\text{PS}_k)$ is the unknown total length of the k-th PS lines that will be eliminated from the list of the unknown variables by dividing member to member the Equations (2)–(4) by Equation (1). In this way, the system will be reduced to three equations with three unknown variables ($nF(\text{U})$, $nF(\text{R})$, and $nF(\text{I})$). On the contrary, the lengths and the energy consumptions of the representative feeders are fixed and defined by Tables 2 and 3.

It is important to observe that if the portion of territory supplied by a given PS is totally rural, only typical rural feeders will be used for modelling the grid, or, if one ambit is missing, the model takes into consideration only the consistent ambits (e.g., if there is no part of territory classified as industrial, i.e., $I_\% = 0$, the number of industrial typical feeder $nF(I)$ is zero according to Equation (3)).

The number of feeders of each representative type, resulting from solving the system of Equations (1)–(5), is generally not an integer. The final result is obtained by minimizing the difference between the total energy required by the loads supplied by the PS and the energy required by each representative feeder (i.e., the difference between the left-hand and the right-hand sides of Equation (5)). It is important to observe that the procedure may add only feeders that match with the land cover and usage of the territory of a given PS. For instance, it is not allowed to add an industrial feeder to a model of a PS that serves a territory without industrial buildings/factories.

At this step, each PS has been modeled in terms of topology and demand consumptions, as a passive PS. It is worth noticing that the consumption profile of the modeled PS may not exactly coincide with the consumption profile attributed to the real PS with the methodology described in Section 3. The residual energy that remains not covered by any feeder of the synthetic network is attributed to a fictitious load/generator, with suitable daily profile, directly connected to the MV busbar of the HV/MV transformer at the PS.

Finally, one or more DG scenarios have to be applied to the model. The DG scenarios of the network model may be constituted by diverse penetration levels and technology mix, including different combinations of non-programmable RES and programmable energy sources (i.e., CHP or biomass generators). One option is to reproduce the realistic scenario of DG derived from step 5 described in Section 3 and allocate generators in the representative feeders that compose the passive model of the PS until it achieves the supposed real DG penetration. Alternatively, future and foreseen scenarios may be hypothesized for planning studies with long or brief term horizons. In the proposed approach, in the default mode, the DG units are chosen among discrete sizes, previously fixed, typical for each technology (e.g., PV sizes can be from 250 kW, while the wind turbines considered range starts from smaller sizes, such as 20 kW or 60 kW). Otherwise, in the case of known power plants, the model can be forced to install DG units that are not of the pre-defined sizes. Once one scenario of DG, real or hypothesized, has been supposed, it has been approximated by minimizing the difference between the supposed installed power and the DG units that will be installed in the model, by combining the pre-defined sizes and the known DG units. The minimization has been performed with a genetic algorithm (GA), where the variables are the number and the size of generators for matching with a given DG penetration. The optimization receives as input the number of typical feeders, candidates for positioning a certain quantity of DG, and provides which feeders of the model the DG units will be connected and of which size as output. The deviation between the supposed DG installed power and the PS model is the objective function to be minimized by the GA. Such deviation depends on the discretization that has been chosen in the definition of the DG technologies sizes (for each technology, the difference can be at maximum equal to the minimum size) and is compensated with the above-mentioned fictitious feeder introduced for adjusting the load demand.

Furthermore, the share of technologies and the specific position of the DG units along the typical feeders can be randomly distributed in the network model or, alternatively, some inputs to better modelling the real condition may be used for forcing defined positions. In particular, one can choose the share of DG installed in each type of representative feeder, the relative position along the feeder, and also if the scenario considers big generators concentrated in a few nodes of the modelled network, or many small generators distributed in as many nodes. In particular, the possible positions of DG units along the feeders can be:

- At the beginning: close to the HV/MV transformer;
- In the middle: in the central part of the feeder;
- At the end: far from the HV/MV transformer.

The procedure also allows us to fix some positions according to specific information about the real installation position in the network.

It is worth noticing that the differences between the provided inputs affect the resulting scenario that might be more or less critical depending on the undertaken choices (i.e., concentrated big generators cause more operation issues than small and distributed units).

The result of this procedure is the synthetic network of the real network downstream from a given PS, that can be used for all of the kinds of studies that need to check the grid limitations. In the following Section 5.1, an example of application of the described approach is reported.

5. Italian Case Study

The proposed methodology has been applied to six Italian regions that represent three of the six NTG zones on which the Italian territory is subdivided: the two biggest islands, Sicily and Sardinia, that constitute one NTG zone each (SARD and SICI respectively), and the regions Puglia, Basilicata, Molise, and Calabria, in the south-east of Italy, that are grouped in the NTG zone named SUD (Figure 13). Molise and Basilicata are respectively, the second and the third smallest regions in Italy, after Valle d'Aosta in the north Italy. Calabria and Puglia can be identified as the toe and the heel of the Italian

boot, respectively. The total number of examined NTG nodes is almost 800 and more than half (i.e., 477) are public PSs.

Figure 13. Subdivision of the Italian territory in NTG zones [19].

In the considered Italian regions, there is a very high concentration of RES both at transmission and distribution levels (about 30% of summer production is from RES). In some cases, as in Sardinia, other trends may be detected, as decreasing demand, transfer of consumption from the transmission to the distribution systems, mainly due to a progressive de-industrialization. For these trends, detailed grid models for planning purposes, but also for estimating how the distribution systems can contribute to system services, are crucial in such regions. In this part of Italy, the services that could be provided by distribution may be mainly for balancing, to cope with high shares of non-programmable generation.

Firstly, the modelling started by assessing the state of the network at the starting time. Thus, the survey of the HV nodes of the six region power systems and their subdivision into the US, PP, and distribution PS has been preventively performed. In Table 4, the share of the different types of NTG nodes for each zone has been reported. Other means substations types different from PP, PS, and US (e.g., railway, auxiliary, hydro pumping, etc.).

**Table 4.** Share of the different types of National Transmission Grid (NTG) nodes.

NTG Zone	PP	PS	US	Other	Total
SUD	58	261	30	54	403
SARD	42	79	22	9	152
SICI	13	137	17	50	217

In Table 5, the total demand and the share of customers categories for each region, as made available by Reference [18], for the year 2015, is reported. By applying the above-described top-down procedure, twelve daily consumption profiles, representative of typical days of the year, have been obtained for each PS relevant to each administrative region.

In order to clarify the details of the methodology, let us consider the Sardinia Island, which is characterized by an area of 24,090 km^2, sparsely populated by 1.6 million inhabitants, supplied by 152 HV/MV primary substations, 79 of which are public PSs. The smallest portion of territory used for the territorial segmentation is the municipality. In Figure 14, such territorial segmentation of the region and the public PSs position are shown.

Table 5. Share of energy consumptions of the examined Italian regions.

NTG Zone	Region	Agricultural cons. (%)	Industrial cons. (%)	Tertiary cons. (%)	Residential cons. (%)	Total Energy (GWh)
SUD	Basilicata	2%	54%	24%	20%	2548.9
	Calabria	3%	15%	43%	40%	5140.8
	Molise	3%	46%	29%	22%	1284.8
	Puglia	3%	44%	28%	25%	16479.8
SARD	Sardinia	3%	45%	27%	26%	8403.46
SICI	Sicily	2%	33%	32%	33%	17214.7
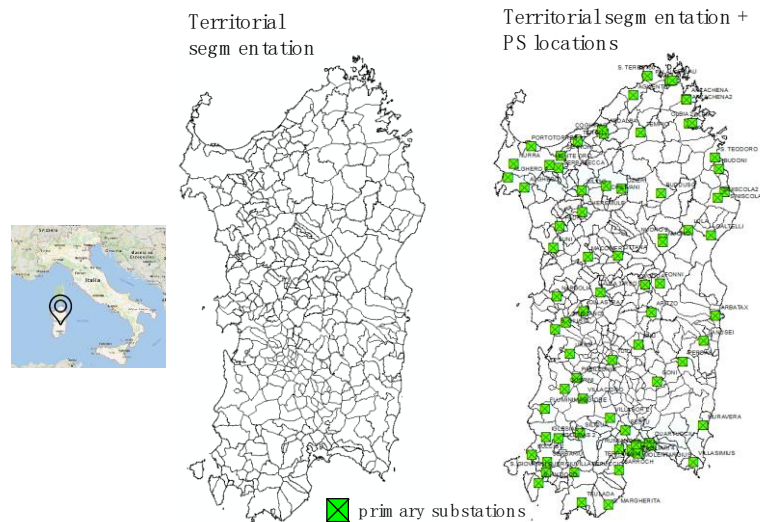
Figure 14. The territorial segmentation of the region and the public PSs position.

Starting from this segmentation, by resorting GIS applications and tools with the information gathered from regional and administrative bodies, a high spatial resolution classification has been made: each portion of territory has been associated with the related population, number of employees, number of buildings, land cover, and land usage according to the procedure described in Section 3. Then, by spreading the total regional load demand throughout the eight Sardinian provinces and the provincial demand throughout the municipalities that belong to each province, the share of demand consumption has been assigned to each territorial portion.

The most difficult task is the association with each PS to the supplied portion of the territory. In the proposed approach, this task is solved by the development of a semi-automatic procedure described in Section 3. By resorting to an iterative process that converges when at least a small portion of each municipality is covered by circles of influence, built according to the rules mentioned above, the procedure allows for defining an incidence matrix that associates the territorial portions with one or more PSs. In particular, in Figure 1, the steps of the procedure applied to one of the provinces of Sardinia, the most southern one, where the capital of the region is located, the city of Cagliari is shown. The Province of Cagliari is subdivided into 71 municipalities (about 4570 km^2), supplied by 19 public PSs. The other steps of the procedure proceed by exploiting the land usage information. In the considered territory, the land usage may be classified as predominantly rural and only close to Cagliari and its hinterland does it change to urban and industrial classifications (Figure 15).
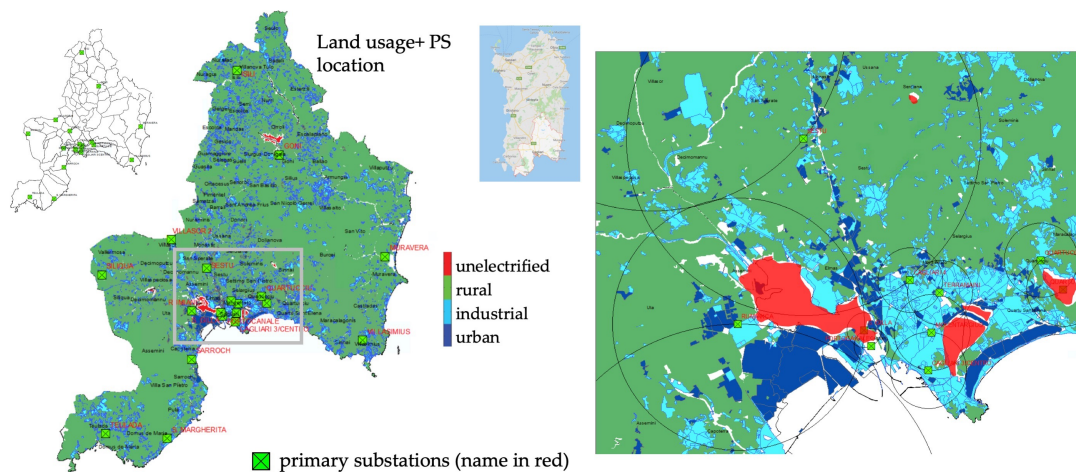
Figure 15. Spatial downscaling: land usage and PS location of the province of Cagliari. On the right, the zoom of the area close to Cagliari (highlighted with the grey rectangle on the left). In the background, the circles used for building the impedance matrix are also visible.

Finally, according to the consumption category percentage obtained with the spatial classification, and by combining the daily reference load profiles of the different categories of customers (Figures 6, 8 and 10), the consumption profiles of each HV node of the whole Sardinian power system have been assessed. According to the information about the installed power generation provided by Reference [24], these profiles have been adjusted by considering the production of each PS (in Sardinia, PV generators installed at the distribution level are predominant). The resulting profiles represent, for each time interval and for each PS, the expected exchanged power through the HV/MV transformers, that represent the TSO/DSO interfaces. It is worth noticing that economic, social, and demographic characteristics and RES availability strongly impact on the shape of load/generation curves and the energy quantities, that obviously follow the law: the greater urbanization the higher the consumption. For instance, in Figure 16, the resulting exchange profiles of two PSs, one in an urban area and one in one predominantly rural area, with a high share of DG. The first is characterized by high demand (positive), while the second shows power flow inversions (negative) through the HV/MV transformer, mainly during spring and summer, due to the high quantity of PV installed.
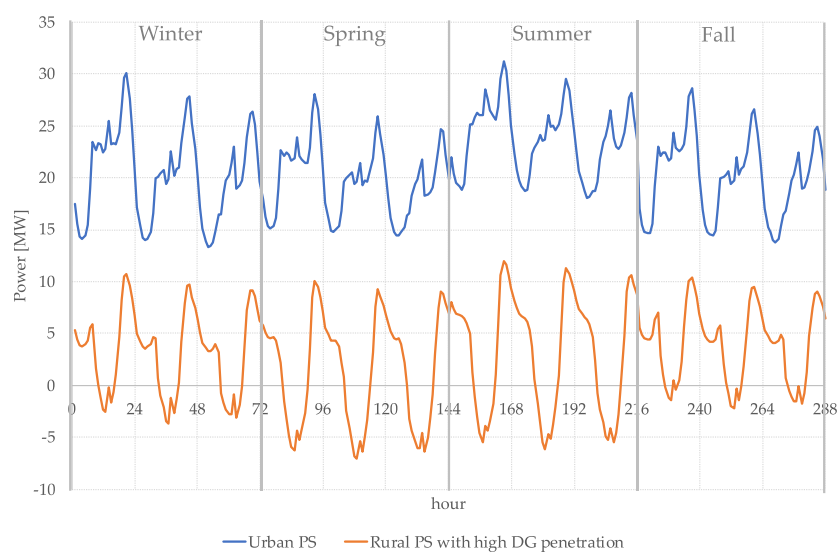

Figure 16. Power exchange with the HV grid (balance of demand and production) of two PSs: one located in urban ambit and one predominantly rural with a high share of DG.

Besides the attribution of the demand and production profiles to each PS, the other main task of the proposed methodology is modelling the real networks' downstream PS with synthetic networks, by using a suitable combination of the typical feeders described in Tables 2 and 3.

Table 6 shows the results of the application of the procedure described in Section 4 to the 79 PSs in the considered region. In the table the number of representative feeders that constitute the synthetic networks are detailed (where R is for rural, U for urban, and I for industrial typical feeder) and also the accuracy of the representation ε_{model} (i.e., the percent difference between the energy demanded by the loads of the PS model and the estimated true PS energy demand, ED$_{PS}$ (GWh/y)). It is worth noticing that the proposed modelling of the whole Sardinian distribution system goes wrong for only 2% of the total energy demand and 85% of the PS has been modeled with an accuracy less than 20%.

5.1. Single Network Results

From the procedure described in Section 3, the PS No. 52 of Table 6, considered for this example, feeds customers that consume 81.4 GWh/year. The resulting share of ambits, derived from the GIS application, by intersecting shapefiles of buildings, unelectrified areas, and land usage of the served territory, is 92.5% of rural, 6.2% of urban, and 1.3% of industrial ambit, respectively. In Figure 17, the PS synthetic network is shown. The model provided by the proposed approach is constituted by six feeders, two of rural ambit (feeder R1 and R2), one dedicated feeder for the biggest PV and CHP power plants (feeder D3), and three urban feeders (feeder U4–U6), two of which are passive feeders. The total installed DG capacity power is about 26 MW of PV, 2.5 MW of WIND, and 5.2 MW of CHP. In this case, the DG position is chosen and distributed at the end of only one of the rural feeders for the PVs, and at the end of the two rural feeders for the WIND turbines, while the CHP is considered concentrated at the beginning of one rural and one urban feeder. Furthermore, the dedicated feeder includes two known existing plants, one 9.6 MW PV and one 2.7 MW CHP. The difference between the annual consumption of the real network and the model is about 8.6%, while the deviation from the hypothesized DG scenario is about 1.3 %.
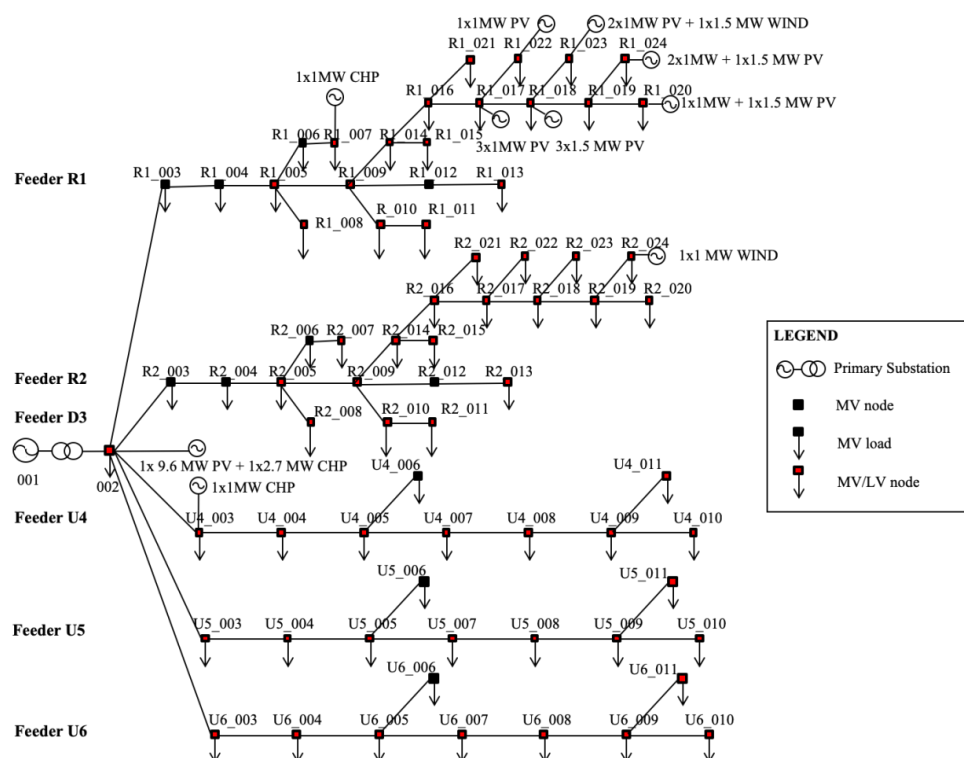

Figure 17. Example of the synthetic network (model of the PS N. 52 of Table 6).

Table 6. Synthetic network composition of the distribution network supplied by the 79 public PS of Sardinia island (ED_{PS} is the PS annual energy demand, ε_{model} is the accuracy of the representation, and R, I and U are used for rural, industrial and urban typical feeder, respectively).

PS No.	Number of Feeders	ED_{PS} (GWh/y)	ε_{model} (%)	PS No.	Number of Feeders	ED_{PS} (GWh/y)	ε_{model} (%)	PS No.	Number of Feeders	ED_{PS} (GWh/y)	ε_{model} (%)	PS No.	Number of Feeders	ED_{PS} (GWh/y)	ε_{model} (%)
1	4 U	66.6	−19%	21	1 R; 11 U	176.6	−6%	41 *	1 R; 3 U	58.8	−2%	61	1 R; 6 U	108.1	−9%
2	2 U	32.7	−17%	22	1 R; 5 U	79.9	6%	42 *	5 R; 15 U	294.3	−2%	62	1 R; 2 U	52	−5%
3	3 U	43.3	−6%	23	1 R; 2 U	32.3	36%	43	1 R; 3 U	59.6	−4%	63	3 U	49	−17%
4	2 U	23.8	14%	24	2 R; 3 U	70.4	6%	44	2 U	29.7	−9%	64	2 U	25.7	5%
5	1 R; 3 U	58	−1%	25	1 R; 4 U	78.4	−9%	45	1 R; 3 U	69	−17%	65	4 U	57.9	−6%
6 *	2 R; 1 U	53.7	−12%	26	2 R; 3 U	71.9	3%	46	1 R; 2 U	35	26%	66	1 R; 2 U	29.4	50%
7	3 R; 6 U	127.3	4%	27	1.U	8.3	63%	47	1 R; 4 U	73	−3%	67	2 U	24.4	11%
8	3 U	50.7	−20%	28 *	2 U	38.6	−30%	48 *	1 R; 3 U	49.6	16%	68	1 R; 4 U	70.3	1%
9	1 U	24.1	−44%	29 *	3 U	41.6	−3%	49	3 R; 11 U	212.7	−6%	69 *	1 R; 4 U	82.7	−14%
10	1 R; 3 U	62.6	−8%	30 *	3 U	44.5	−9%	50 *	2 R; 9 U	171.3	1%	70 *	2 R; 4 U	92.3	−5%
11	3 U	39	4%	31	2 U	31.2	−13%	51	1 R; 2 U	44.8	−2%	71	1 R; 2 U	36.3	21%
12	1 U	19.4	−30%	32 *	4 U	61	−11%	52 *	2 R; 3 U	81.4	−9%	72	1 R; 1 I; 10 U	171.7	1%
13	1 I; 3 U	59.1	5%	33	1 R; 8 U	137.5	−9%	53 *	1 R; 4 U	63.4	12%	73	2 U	38.3	−29%
14 *	1 R; 1 I; 3 U	77.6	2%	34 *	3 R; 1 I; 8 U	181.4	−1%	54 *	1 R; 5 U	84.1	0%	74	2 R; 5 U	104.8	−3%
15 *	1 R; 5 U	83.8	1%	35	1 R; 13 U	203.6	3%	55 *	1 R; 3 U	51.4	12%	75	1 R; 10 U	160.9	−6%
16	2 R; 1 I; 7 U	143	5%	36	2 R; 4 U	86.9	2%	56	3 R; 6 U	128.9	2%	76	1 U	15	−9%
17	1 R; 8 U	129.1	−3%	37	2 U	36.3	−26%	57	1 I; 4 U	72.4	4%	77	1 R; 5 U	87.1	−3%
18 *	2 U	25.4	7%	38 *	4 R; 1 I; 11 U	227.3	5%	58	12 U	173.1	−6%	78 *	3 R; 7 U	141.6	3%
19 *	1 U	25.9	−48%	39 *	2 R; 2 U	60.5	1%	59 *	14 U	201.6	−6%	79	10 U	139.9	−3%
20 *	4 R; 5 U	132.3	2%	40	2 U	29.9	−10%	60 *	2 R; 9 U	153.4	1%				

* with a further dedicated feeder for connecting big DG plants.

6. Conclusions

The paper proposed a novel methodology that allows for modelling the distribution systems at a regional/zonal scale by using only open data sources. The original procedure was able to model the distribution networks connected to a portion of the bulk grid starting from social and economic information regarding the territory served by each primary substation and by combining a suitable set of representative network portions. This modular approach permits us to overcome the issues related to the distribution system dimension and diverse contexts. The result is the synthetic network of the actual distribution network that can be used for all the kinds of optimization studies that need to check the grid limitations. The methodology has been tested in six Italian regions.

References

1. IEEE PES AMPS DSAS Test Feeder Working Group, Test feeders. Available online: http://sites.ieee.org/pes-testfeeders/resources/ (accessed on 22 November 2019).
2. Strunz, K.; Abbasi, E.; Fletcher, R.; HatziargyriouOptimal, N.D. *TF C6.04.02: TB 575—Benchmark Systems for Network Integration of Renewable and Distributed Energy Resources*; CIGRE: Paris, France, 2014.
3. Pilo, F.; Pisano, G.; Soma, G.G. Optimal Coordination of Energy Resources with a Two-stage on-line Active Management. *IEEE Trans. Industrial Electron.* **2011**, *58*, 10. [CrossRef]
4. Di Fazio, A.R.; Russo, M.; De Santis, M. Zoning Evaluation for Voltage Optimization in Distribution Networks with Distributed Energy Resources. *Energies* **2019**, *12*, 390. [CrossRef]
5. Seritan, G.; Porumb, R.; Cepisca, C.; Grigorescu, S.D. Integration of Dispersed Power Generation. In *Electricity Distribution-Intelligent Solutions for Electricity Transmission and Distribution Networks*; Karampelas, K., Ekonomu, L., Eds.; Energy Systems, Springer: Berlin/Heidelberg, Germany, 2016; pp. 27–61.
6. Muñoz-Delgado, G.; Contreras, J.; Arroyo, J.M. Distribution Network Expansion Planning with an Explicit Formulation for Reliability Assessment. *IEEE Tran. Power Syst.* **2018**, *33*, 2583–2596. [CrossRef]
7. Ceaki, O.; Vatu, R.; Golovanov, N.; Porumb, R.; Seritan, G. Analysis of the Grid-Connected PV Plants Behavior with FACTS Influence. In Proceedings of the 49th International Universities Power Engineering Conference (UPEC 2014), Cluj-Napoca, Romania, 2–5 September 2014.
8. Nick, M.; Cherkaoui, R.; Paolone, M. Optimal Planning of Distributed Energy Storage Systems in Active Distribution Networks Embedding Grid Reconfiguration. *IEEE Trans. Power Syst.* **2018**, *33*, 1577–1590. [CrossRef]
9. Bracale, A.; Caldon, R.; Celli, G.; Coppo, M.; Dal Canto, D.; Langella, R.; Petretto, G.; Pilo, F.; Pisano, G.; Daniela, P. Analysis of the Italian Distribution System Evolution through Reference Networks. In Proceedings of the 2012 IEEE PES Innovative Smart Grid Technologies (ISGT) Europe Conference, Berlin, Germany, 14–17 October 2012.
10. United Kingdom Generic Distribution Network (UKGDS). Available online: https://github.com/sedg/ukgds (accessed on 22 November 2019).
11. Prettico, G.; Flammini, M.G.; Andreadou, N.; Vitiello, S.; Fulli, G.; Masera, M. *Distribution System Operators Observatory 2018 Overview of the Electricity Distribution System in Europe*; Publications Office of the European Union: Luxemburg, 2019.

12. Fulli, G.; Mateo, C.; Prettico, G.; Gómez, T.; Cossent, R.; Gangale, F.; Frías, P.; Fulli, G. European representative electricity distribution networks. *Elsevier Electr. Power Energy Syst.* **2018**, *99*, 273–280.

13. Liao, Y.; Weng, Y.; Liu, G.; Rajagopal, R. Urban MV and LV Distribution Grid Topology Estimation via Group Lasso. *IEEE Trans. Power Syst.* **2019**, *34*, 12–27. [CrossRef]

14. Duan, Y.; Wang, C.; Zhou, W. Topology modeling of distribution network based on open-source GIS. In Proceedings of the 2011 4th International Conference on Electric Utility Deregulation and Restructuring and Power Technologies (DRPT), Weihai, Shandong, China, 6–9 July 2011.

15. European Commission. Clean Energy for all Europeans Package (Nov. 2016). Available online: https://ec.europa.eu/energy/en/topics/energy-strategy-and-energy-union/clean-energy-all-europeans (accessed on 22 November 2019).

16. Italian Regulatory Authority for Electricity Gas and Water. Documento per la Consultazione 354/2013/r/eel. Available online: http://www.autorita.energia.it/allegati/docs/13/354-13all.pdf (accessed on 22 November 2019). (In Italian).

17. Italian Regulatory Authority for Energy, Networks and Environment, Deliberazione 5 maggio 2017 300/2017/R/EEL. Available online: https://www.arera.it/allegati/docs/17/300-17ti.pdf (accessed on 22 November 2019). (In Italian).

18. Terna, SPA. (Italian TSO). Available online: http://www.terna.it/ (accessed on 22 November 2019).

19. Italian National Energy Market Operator (GME). Available online: http://www.mercatoelettrico.org/ (accessed on 22 November 2019).

20. Italian National Geoportal. Available online: http://www.pcn.minambiente.it/mattm/en/ (accessed on 22 November 2019).

21. European Environment Agency, Corine Database. Available online: https://www.eea.europa.eu/publications/COR0-landcover (accessed on 22 November 2019).

22. Italian Institute for Environmental Protection and Research. Corine Land Cover 2012. Available online: http://www.isprambiente.gov.it/it/temi/biodiversita/documenti/corine-land-cover-clc (accessed on 22 November 2019). (In Italian).

23. Italian National Statistics Institute (ISTAT). Available online: http://www.istat.it/ (accessed on 22 November 2019).

24. Italian National Energy Service Manager (GSE). Available online: http://www.gse.it/ (accessed on 22 November 2019).

25. Italian DPR n. 412. 1993. Available online: https://www.gazzettaufficiale.it/eli/id/1993/10/14/093G0451/sg (accessed on 22 November 2019).

26. Cantù, M.; Coppo, M.; Gigliucci, G. Representative Distribution Network Models for Assessing the Role of Active Distribution Systems in Bulk Ancillary Services Markets. In Proceedings of the 19th Power Systems Computation Conference (PSCC), Genoa, Italy, 20–24 June 2016.

27. Gómez, T.; Mateo, C.; Sánchez, Á.; Frías, P.; Cossent, R. Reference network models: A computational tool for planning and designing large-scale smart electricity distribution grids. In *High-Performance Computing in Power and Energy Systems*; Kumar, S., Gupta, A., Aluru, S., Gopalakrishnan, K., Eds.; Springer Verlag: Berlin/Heidelberg, Germany, 2012.

28. Levi, V.; Strbac, G.; Allan, R.N. Assessment of performance-driven investment strategies of distribution systems using reference networks. In *IEE Proceedings Generation Transmission and Distribution*; IET: London, UK, 2005; Volume 152, pp. 1–10.