Data Analytics and Stochastic Simulation Methods for Risk-Controlled Network Planning: Validation Case Study

André ÁGUAS, EDP Distribuição – Portugal andre.aguas@edp.pt

Vera PEREIRA EDP Distribuição – Portugal vera.pereira@edp.pt

Inês ROÇA EDP Distribuição – Portugal ines.roca@edp.pt

Luísa JORGE EDP Distribuição – Portugal luisa.jorge@edp.pt

Ricardo PRATA EDP Distribuição – Portugal ricardo.prata@edp.pt

João MACHADO AmberTREE – Portugal joao.machado@ambertree.pt

Pedro CARVALHO AmberTREE – Portugal pedro.carvalho@ambertree.pt

Luís FERREIRA AmberTREE – Portugal marcelino.ferreira@ambertree.pt

ABSTRACT

EDP Distribuição is developing initiatives aimed at adapting to new realities and technologies, taking advantage of the continued technological investments being made in AMI. In this context, specific data analytics methodologies were developed to characterize the stochasticity of real data from measured load profiles and to adapt planning tools to such paradigm change. This paper presents a comparison between real-time measurements data and synthetic load and power-flow results, in order to validate the methodologies developed.

INTRODUCTION

In recent years, EDP Distribuição (EDPD) started a massive deployment of metering infrastructures and a large volume of metering data became available. The proper use of such data is key to enhance the support to investment decisions, increasing responsiveness and quality of service, while controlling the associated risks [1, 2].

In the current context of changes, we have developed a specific data analytics methodology (detailed in [3]) to explore the large volume of metering data aiming at clustering customer profiles into typical load/generation profiles. The stochastic properties of real data from measured profiles (load and generation) were then synthesized into a set of parameters which reflect load dynamics realistically [1, 4]. Such understanding of load and generation patterns was used to enhance the support to investment decisions by embracing an explicit risk-controlled probabilistic decision-making paradigm [1].

As this novel approach to network planning is significantly different from the traditional methods and involves a series of steps of data analysis, it is exceedingly important to validate the results of the simulations by comparing them to real data. This paper intends to assess those results, by presenting a comparison between the calculated synthetic profiles and the metered profiles, for a specific case study.

LOAD BEHAVIOUR CHARACTERIZATION

One of the main challenges in planning electricity distribution systems is dealing with the uncertainty of load and generation. This uncertainty makes the dynamics of consumption and production very difficult to model, since a good characterization of their behaviour requires a deep knowledge of the correlations and time-dependencies, simultaneously [4]. Taking advantage of the large volume of historical metering data, EDPD has been participating in the development of new data analytics aimed at (*i*) clustering the universe of primary substations, secondary substations, clients and producers' loads uses into typical load/generation profiles and (*ii*) characterizing the corresponding behaviour patterns (dynamics) of such load uses as non-stationary stochastic processes [1, 3]. Figure 1 shows a disaggregation of the developments undertaken.

Clustering

Encouraged by an initial data exploration that revealed different behaviours according to the season of the year and day of the week (business day, Saturday and Sunday), the metered data was aggregated into twelve average profiles for each site after removing outliers.

Given the volume of data considered in this process, as well as the quality of information required, we used data mining techniques such as density-based clustering for data validation as well as hierarchical clustering methods to group the profiles according to the similarities between consumption patterns. At the end, each load has been characterized by twelve daily profiles chosen from the corresponding cluster centroid together with their corresponding mean and standard deviation moments [1].

Furthermore, to assign a typical load/generation profile to future installations or to overcome issues related with metering errors or missing data, an algorithm based on Decision Trees was also developed. This algorithm can classify a load profile with missing information based on the previous classified sites and their corresponding technical and commercial features.

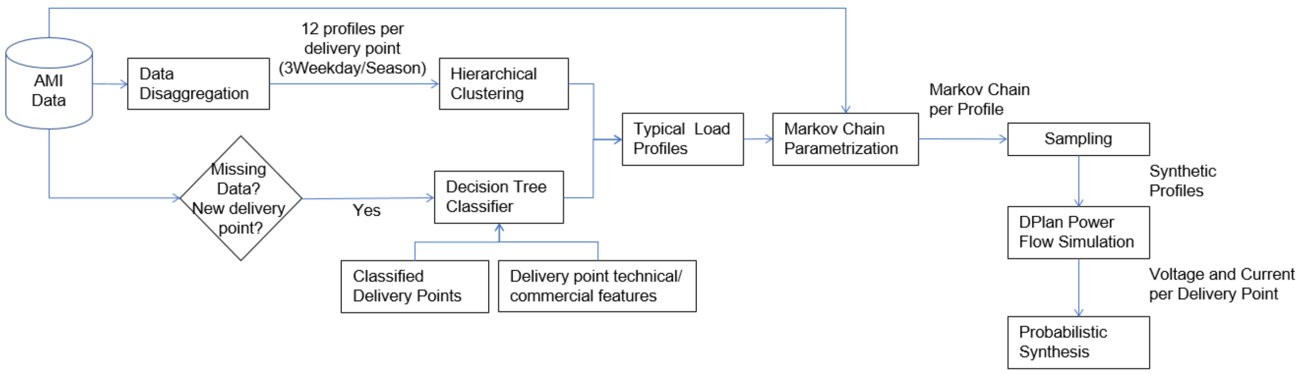


Figure 1 – Disaggregation of developments in data analytics and stochastic simulation.

Behaviour Modeling

The cluster centroids and AMI data were used to model load dynamics for different loads through a stochastic Markov process. The method relies upon generic individual measured profiles to create a discrete-time stationary Markov process to parameterize the average volatility and time-dependency in the corresponding group. The measured profiles of active power are discretized into a set of representative power states. The individual profiles allow to determine the probability of the consumer being in a given load state as well as the expected power in that state. Furthermore, it is possible to define a Markov matrix $A = [a_{ij}]$ that represents the probability of a consumer being in load state i, and change to load state j. Despite the Markov process defined by A may correctly parameterize the average power dynamics of an average consumption or generation profile, it neglects that the probability of being in each load state changes throughout the day. Therefore, transition matrices must change in order to ensure that average simulated power in each time period evolves according to the aggregate profile of the cluster [4].
As a result, the dynamics of a given type of consumer is parameterized into a set of Markov transition matrices, that change over time and synthesize the probability of being in load state i at time t, and change to state j at time $t + 1$, according to the times of the day and the time resolution. Since the real-time measurements have a time resolution of 15min, a chain of 95 Markov transition matrices were modelled for each cluster and type of distribution network sites, where each one has 25 representative load states.

SIMULATIONS AND RESULTS

A comparison between the synthetic profiles generated through the Markov behaviour modelling summarized above and the metered profiles has been carried out. We used an existing LV grid (topology depicted in Figure 2), as well as the load data of such grid, metered between Jan and Mar 2017 (data collected in the framework of the Sensible EU project [5]).
In order to evaluate the accuracy of the representation of load between voltage levels, three simulations were performed for:

1. A specific secondary substation (SS) load
2. A specific low voltage (LV) client load
3. A LV feeder that serves a group of household LV clients.

Additionally, we also analysed the power-flow results obtained for the specific LV network from which AMI data was gathered.

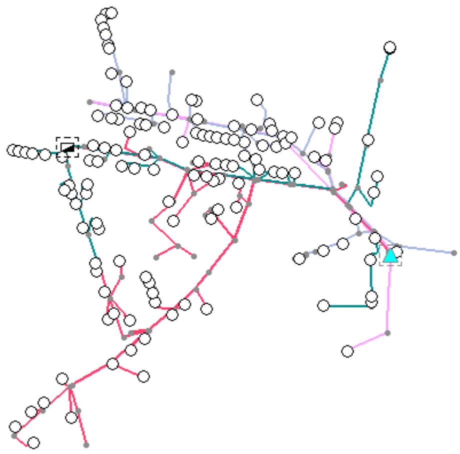
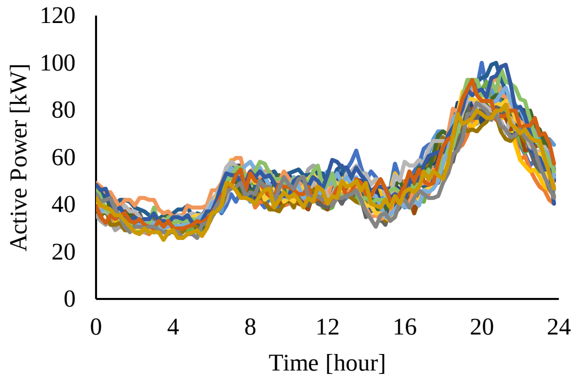


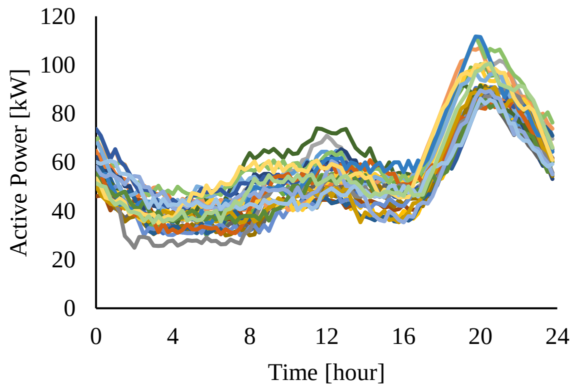
Figure 2 - Schematic representation of the LV grid.

Individual Load Profiles

The first simulation presented focussed on assessing the accuracy of the characterization of one SS load. The following figures show (*i*) a sample of real profiles measured at the SS and (*ii*) a sample of synthetic profiles generated by the Markov process. The juxtaposition of both samples allows us to observe the similar behaviour between the synthetic and the real load profiles. From the figures, it is also apparent that there are slight differences in behaviour and standard deviation. These differences can be explained by the deviation of this particular SS from the cluster it was included in.



a) Real profiles



b) Synthetic profiles

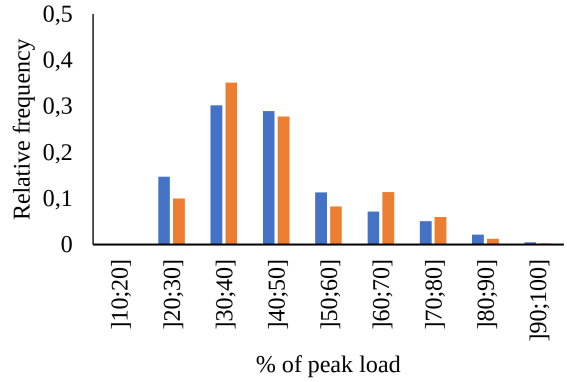
Figure 3 - Real Profiles (a) and Synthetic Profiles (b) for a SS. The synthetic profiles shown are the results from 30 independent simulations. The real profiles shown correspond to all business days of January 2017.

As the synthetic load profiles generated have to represent all loads in this cluster, they usually have a wider variety of values in each time period. In this particular example, we can see that the synthetic peak load can be slightly superior to the real peak load. It can also be observed that until 16:00, there is a wider standard deviation in the synthetic load than in the measured profiles.

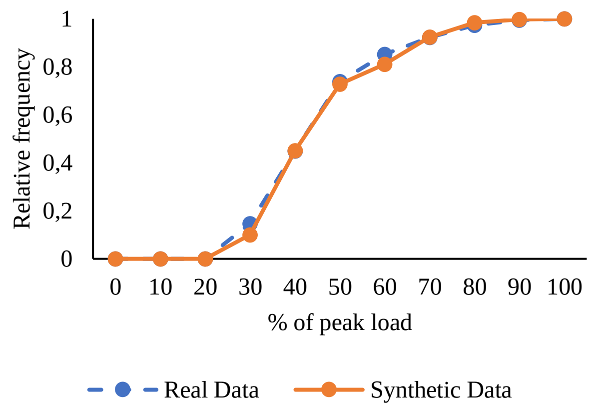
Despite the wider range of load possible values in each time period, the load distribution and dynamics are seemingly well characterized. To better compare the load distributions of each sample, a cumulative distribution and the corresponding histogram are shown (Figure 4).

As can be seen, in both the histogram and the cumulative distribution, the synthetic and real data are very similar. The similarity of both distributions indicates that the load is well characterized and can be used efficiently in distribution planning studies.

The accuracy of each individual load characterization is dependent on the load volatility as well as on its overall distance to the cluster's centroid. As such, LV clients are especially hard to model.



a) Histogram



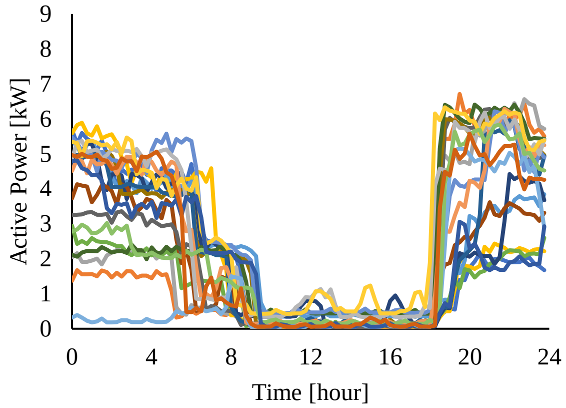
b) Cumulative Distribution

Figure 4 - Histogram for the load distribution of SS (a) and corresponding cumulative distribution (b) of both real load profiles (blue) and synthetic load profiles (orange).

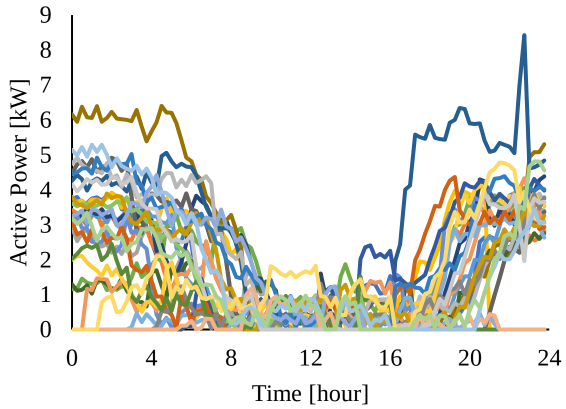
The second simulation focussed on assessing the accuracy of the characterization of one LV client. In Figure 5, the synthetic profiles can be compared with the real daily profiles of an individual LV client.

As expected, individual load characterization proved to be more challenging due to the increased volatility of the LV clients and wider range of possible load values in each time period analysed. However, despite these difficulties, the synthetic load profiles show recognizable consumption behaviours. Both the measured and synthetic daily load profiles show reduced power consumption during the afternoon period and sharp load increases around 18:00. From Figure 5 it can also be observed that these sharp load increases in the synthetic profiles can start earlier or later than in the real data and have higher or lower peaks, depending on the simulation. This can be explained by the different behaviours of clients included in the same cluster.

In Figure 6 we show the histogram and cumulative distribution of both samples. It is important to notice that the synthetic data has very similar average consumption and congruent load distributions which make it suitable to perform a wide range of simulations.

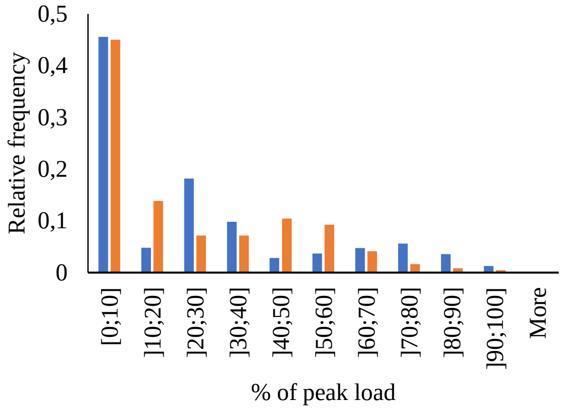


a) Real profiles

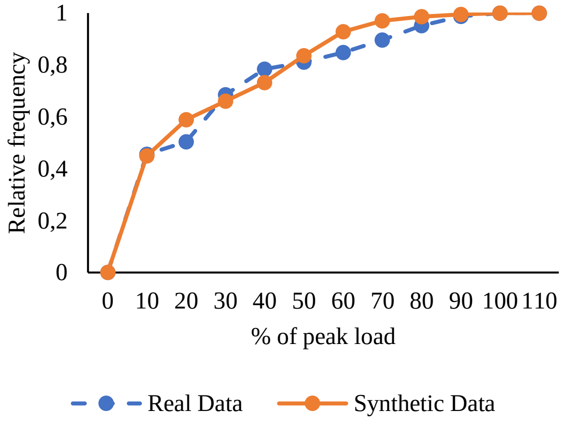


b) Synthetic profiles

Figure 5 - Real Profiles (a) and Synthetic Profiles (b) for a LV Client. The synthetic profiles shown are the results from 30 independent simulations. The real profiles shown correspond to all business days of January 2017.



a) Histogram



b) Cumulative Distribution

Figure 6 - Histogram for the load distribution of a LV Client (a) and corresponding cumulative distribution (b) of both real load profiles (blue) and synthetic load profiles (orange). From the cumulative distribution, it can be observed that the average load is almost the same and that the distribution tail is well simulated by the synthetic data.

Aggregate Load Profiles

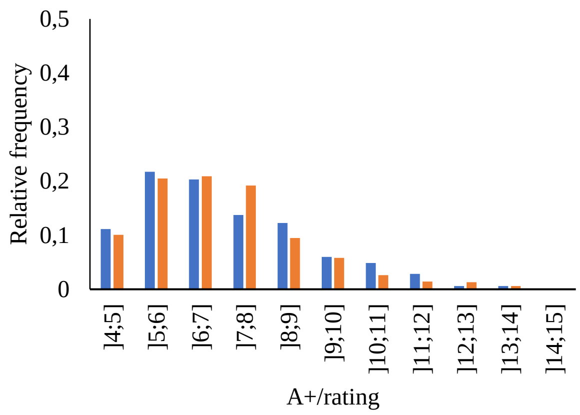
Although individual load characterization suffers from the deviations of each load to their own cluster, when analysing a feeder or a secondary substation these deviations are less significant. This can be explained because each individual client has their own differences from the corresponding cluster characterizations and such differences tend to cancel out when summed.

To test this assumption, a single feeder with 15 LV clients of varying contracted powers (from 1,15 kVA to 20,7 kVA) and different connection phases was chosen from the LV network in Figure 52 (pink line). Despite the different contracted powers, the phase connection imbalance and the small number of connected clients, we have consistently experienced accurate representations of the aggregate load profile. Figure 7 shows the histograms and accumulated distributions of both the synthetic daily load profiles and the real metered load profiles. As can be observed, the histogram shows very similar frequency distributions throughout all the load intervals considered.

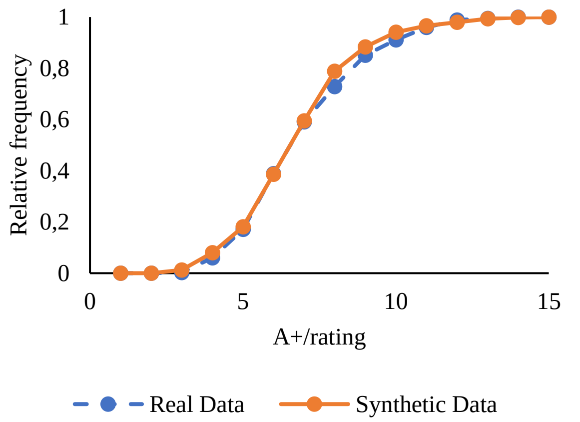
It is also important to notice that, besides showing a good characterization of the average client in each cluster, the feeder aggregation shows that the daily load profiles also correctly characterize daily load dynamics and the correlation of load consumption between LV clients. As we show in the next section, by characterizing the load distribution through daily load profiles, we can correctly evaluate the impacts of correlated but highly volatile loads, allowing us to have significant advantages in, for example, assessing node voltages in volatile and unbalanced networks.

Case Study Results

By using synthetic data to characterize load consumption we can benefit from having each and every load characterized in a way that broadly represents its usual consumption pattern, while avoiding errors from missing data points in real metering data.



a) Histogram



b) Cumulative Distribution

Figure 7 - Histogram for the load distribution of the aggregate load (a) and corresponding cumulative distribution (b) of both real load profiles (blue) and synthetic load profiles (orange).

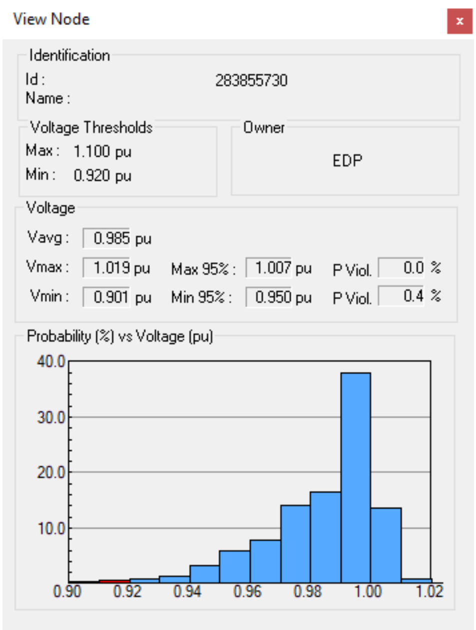


Figure 8 - Voltage histogram of a LV connected client.

Daily load profiles can be used to perform multiple grid analysis simulations and obtain state probable values of the grid currents and voltages for each daily time period evaluated. Using DPlan's probabilistic simulation and synthesis tools, we promptly performed thousands of power flows to evaluate voltage fluctuations and branch currents. For the network of Figure 2, simulations revealed voltage quality issues on specific grid locations and allowed pinpointing the corresponding loading situations, otherwise impossible to identify due to the high load volatility and significant phase imbalance of neighbourhood loads. Figure 8 illustrates the histogram of the nodal voltage at the connection point of one LV client. The figure shows that voltage may drop below 0.92 p.u., despite such possibility being very small.

CONCLUSIONS

This paper compares real metering data with synthetized load profiles results obtained by clustering and characterization with Markov processes. Comparisons are made to evaluate the underlying capability of carrying out reliable planning studies based on synthetized results. Comparisons have shown that individual synthetized load followed the real metering data consumption patterns with acceptable accuracy even for LV consumers whose intrinsic volatility and inherent cluster classification errors were significant. Despite some imperfections in single load characterization, the differences in the load distributions between synthetic and measured data start to quickly fade when aggregating loads or analysing realistic grid loading situations: in a challenging test over a LV feeder, the differences between the synthetic data and the measured data were shown to be negligible and to be reliable in providing information on limiting operational conditions crucial for planners.

REFERENCES

[1] A. Águas, V. Pereira, P. Carvalho, J. Machado, et al., 2018, "EDPD's experience with data analytics and stochastic simulation methods for risk-controlled network planning", *CIRED Ljubljana Workshop*.

[2] P.M.S Carvalho, L.A.F.M Ferreira, A.M.F. Dias, 2018, "Distribution grids of the future: Planning for flexibility to operate under growing uncertainty", *Foundations and Trends in Electric Energy Systems.* vol. 2, 324-415.

[3] V. Pereira, P. Mousinho, L. Jorge, 2017, "Identification of electrical energy consumption patterns", *CIRED 24th Edition*.

[4] J.A.C Machado, P.M.S Carvalho, L.A.F.M Ferreira, 2018, "Building stochastic non-stationary daily load/generation profiles for distribution planning studies", *IEEE T. on Power Systems.* vol. 33, 911-920.

[5] H2020 Sensible Project (Grant Agreement n.º 771066)