EDPD'S EXPERIENCE WITH DATA ANALYTICS AND STOCHASTIC SIMULATION METHODS FOR RISK-CONTROLLED NETWORK PLANNING

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

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

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

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

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

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

Rui BENTO
EDP Distribuição – Portugal
ruimiguel.bento@edp.pt

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

ABSTRACT

Within the framework of the Move2Future project, EDP Distribuição (EDPD) is adapting to new realities and technologies, taking advantage of the investments in AMI. In that effort, EDPD embraced the development of a comprehensive methodology to enhance the support to investment decisions, increasing responsiveness and quality of service, while controlling the associated risks. This involved developing specific data analytics approaches for synthetic load modelling based on real data and adapting planning tools to synthetize grid outcomes of such modelling as risk indices. This paper presents the main challenges related to modelling and simulation, pinpointing key issues in computing resources, algorithm design and results presentation.

INTRODUCTION

Utilities are nowadays dealing with a new reality: new loads and distributed energy resources emerged and large volumes of metering data became available. Metering data allows an improved understanding of load and generation patterns which can be used to enhance the support to investment decisions by embracing an explicit risk-controlled probabilistic decision-making paradigm.

Our work in this field is multifold. In this paper, the focus will be on the developed data analytics necessary to characterize loads stochasticity and on the adaptation of the existing simulation tools necessary to deal with such stochasticity and provide the basis for decision-making in grid investment. This is summarized in the following and illustrated in Figure 1.

1) Develop specific data analytics tools to explore the metering data aiming at segmenting customer profiles into typical load/generation profiles, and characterize such profiles to extract representative behaviours to be modelled as stochastic processes parameterized into Markov chains;
2) Evolve the computational applications that support decision-making (DPlan), to input representative load/generation behaviours as parameters of stochastic processes and simulate such processes to provide probabilistic results that support risk-controlled planning decisions over grids.

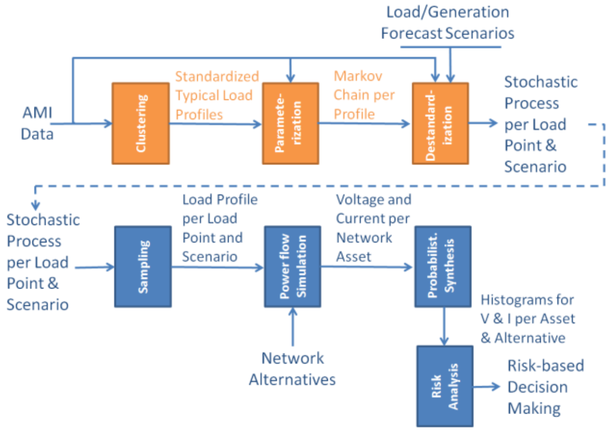

Load/Generation Forecast Scenarios
AMI Data
Clustering
Standardized Typical Load Profiles
Paramete-rization
Markov Chain per Profile
Destandard-ization
Stochastic Process per Load Point & Scenario
Stochastic Process per Load Point & Scenario
Sampling
Load Profile per Load Point and Scenario
Power flow Simulation
Voltage and Current per Network Asset
Probabilist. Synthesis
Histograms for V & I per Asset & Alternative
Network Alternatives
Risk Analysis
Risk-based Decision Making

Figure 1 – Disaggregation of developments in data analytics (above) and stochastic simulation (below) carried out to enable risk-controlled probabilistic decision-making.

METHODOLOGY FOR CLUSTERING DAILY LOAD / GENERATION PROFILES

Encouraged by technological development, utilities are gathering large amounts of AMI data [1]. To deal successfully with this newly available data, both for consumption and for production, EDP Distribuição (EDPD) has been participating in the development of new data analytics. In particular, analytics have been used to cluster time-series of load without human supervision, avoiding preconceived profiling and segmentation.

This section addresses the main steps of the methodology: (*i*) to segment times-series of load data from the universe of distribution network sites (DNS), namely primary substations, secondary substations, clients and producers; (*ii*) to characterize the corresponding behaviour patterns. The segmentation was developed using 'R' programming, which provides adequate statistical computing tools and graphical environments.

An initial analysis showed that, as anticipated, power consumption and production patterns vary according to the season of the year and day of the week (business days, Saturday, Sunday). Therefore, the methodology started by disaggregating the data according to such periods for each DNS.

As there are obvious similarities between consumption patterns within the same period, the second step was to configure a single load profile representative of all days of a given period. To do so, we have identified and eliminated, using density-based clustering methods, the days with abnormal power consumption and calculated the average of the selected remaining profiles [2].

The process of clustering was repeated for the four seasons of the year and the three types of day. As a result, we obtained twelve daily profiles for each DNS, which together characterize DNS annual average behaviour. Figure 2 illustrates the process.

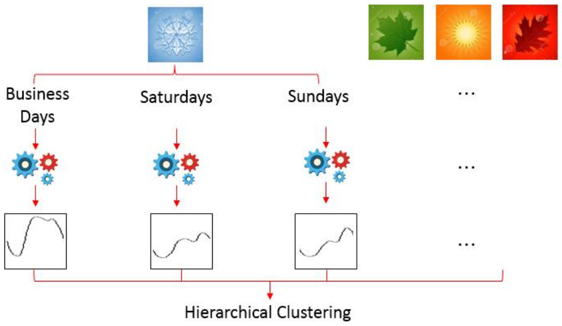


Figure 2 – Summary of the methodology for clustering daily profiles.

The next step of the methodology consisted of grouping DNS time series in clusters according to the power consumption/production patterns. We used hierarchical clustering to accomplish that. As the aim of the methodology was to compare load profiles with each other in terms of shape, we standardized DNS time-series before clustering them. Standardization was done by subtracting the average time series value to the series $[X_k]$ and dividing the result by the standard deviation.

We took different clustering approaches for primary substations (PS) and for secondary substations (SS).

- For PS, the clustering minimized the distances between time series composed by the twelve daily average load profiles of each DNS in a row, resulting in a unique representative load profile to characterize each substation's annual behaviour [2].
- For SS, the previous approach failed to provide good results due to a much higher level of weekly and seasonal load diversity. So, we took the option of clustering daily patterns and then characterizing annual behaviour by the composition of twelve daily representative patterns. Accordingly, the clustering has been set to minimize distances between daily time series, independently[1].

Daily pattern clustering involves a much higher computational effort than the annual pattern approach as the universe of daily time-series is much higher than the number of annual series. In addition to this, the number of SS and LV clients is also much higher than the number of PS, which makes the daily clustering approach computationally expensive. Table 1 shows the total number of DNS for each load category.

Table 1 – Number of DNS for each category.

DNS Category	Number of DNS
Primary Substations	404
Secondary Substations	68 200
Clients (HV and MV)	24 400
Producers (HV and MV)	630
Clients (LV)	6 065 720

After clustering, each DNS is characterized by a sequence of twelve daily profiles chosen from the corresponding cluster centroids. Such sequence of centroid profiles together with the specific DNS load mean and standard deviation (used for profile standardization) are then stored as the minimal necessary information to characterize DNS annual load behaviour individually. Figure 3 illustrates such information for a specific SS case. As it will be explained next, this information will be crucial to set up a discrete-time non-stationary Markov process that realistically reproduces high-resolution daily load volatility and time dependency.

MODELING, SIMULATION AND SYNTHESIS FOR PROBABILISTIC ANALYSIS

Modelling

This section describes how the standardized daily profiles and AMI data are used to model load dynamics and sample load values for each DNS, in each time period, through a stochastic Markov process. The Markov process uses (*i*) the standardized individual load time series clustered under the same pattern to characterize the profile stochasticity and (*ii*) the cluster centroid load profile to characterize the typical intra-day load dynamics of each pattern.

As load state transition probabilities depend on the times of the day, the stochastic process of load is non-stationary. Also, since time resolution of the time series is 15 min, a chain of 95 Markov transition matrices will model daily stochasticity with realistic time-dependencies. Many load states need to be defined in order to discretize load range adequately. We assessed that 25 load states were needed to allow acceptable characterization. The number of Markov chains needed for each daily profile and the size of the state space required to discretize load adequately would make the illustration of the approach taken impractical. Therefore, in his paper, we opt to illustrate the main ideas of the approach with small-scale example.

[1] Proven to be an improvement over the annual pattern approach, the daily pattern clustering is now being used to characterize the annual behaviour of PS as well.

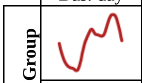
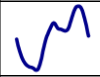
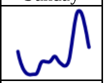
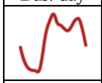
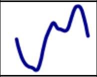
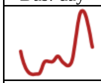
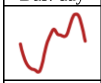
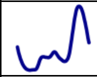
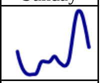
	Winter			Spring			Summer			Autumn		
	Bus. day	Saturday	Sunday	Bus. day	Saturday	Sunday	Bus. day	Saturday	Sunday	Bus. day	Saturday	Sunday
Group												
	3	3	1	4	3	2	1	2	2	3	1	1
μ	74.15	72.58	68.43	58.85	54.40	53.02	53.41	50.58	50.62	58.20	54.16	53.59
σ	14.45	16.41	13.76	10.63	10.09	9.94	8.74	8.20	9.30	11.37	10.94	10.38

Figure 3 – Information needed to characterize a particular SS annual load behaviour: (i) sequence of centroid profiles and (ii) specific DNS load mean and standard deviation. The example highlights the importance of having a methodology which correctly identifies different behaviours over the year. We can see that the typical load profiles of business days are similar in Winter, Spring and Autumn and different in Summer. This results from the fact that this SS feeds a high school, which makes the business day profile more residential-like in the Summer.

Figure 4 illustrates a simple case with a Markov chain with three load states {0, 1 ,2} and the corresponding sequence of transition probabilities as given by the 3×3 matrices, $P(k) = [p_{ij}^k]$, with p_{ij}^k as given below:

$$P(k) = \begin{pmatrix} p_{00}^k & p_{01}^k & p_{02}^k \\ p_{10}^k & p_{11}^k & p_{12}^k \\ p_{20}^k & p_{21}^k & p_{22}^k \end{pmatrix}, k \geq 1.$$

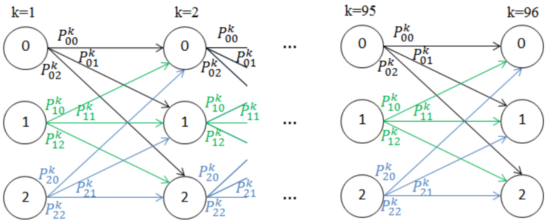


Fig. 4 – Diagram representation of a three-state Markov chain with a sequence of transition probabilities calculated for three states and 96 time periods used to illustrate the representation of intra-day dynamics of daily profiles with 15min resolution.

A more in-depth explanation of the method used to parameterize Markov processes of daily patterns can be found in [3] together with the necessary algorithms to obtain the transition probability matrices.

Once parameterized, the Markov process can be used to sample a sequence of load states [4]. The sampling can be easily done by undertaking a recurrent updating process as follows:

$$X^{k+1} \sim P^{k+1}(X^{k+1} | X^k = X_k)$$
$$X^{k+1} = X_{k+1}$$
$$k \geq 0$$

Where, P^k refers to the probability distribution function represented by matrix $P(k)$. A sequence of load states $[X_k]$ is obtained for each load of the same profile type (same cluster) by sampling. Each new sampling produces a different sequence randomly. Randomness depends only on the profile type. Load specific characteristics (besides profile type) such as loads' expectation and variance are not taken into account in the modelling stage. They will be used only in power-flow simulation.

Before power-flow simulation, sampled sequences $[X_k]$ have to be de-standardized with information on the load specific probabilistic moments. This is done by reverting standardization, i.e., $[X_k] \leftarrow [X_k]\sigma + \mu$.

Simulation

De-standardization is carried out by the grid simulator (DPlan) after a Markov chain being assigned to each i-th bus of the grid. Once assigned representative sequences of load states to each and every bus load, the grid state is simulated by running an AC power-flow for each time period k and year t. Simulation encompasses a whole set of $k \cdot t$ power-flows for each scenario [5]. The power-flow analysis problem can be solved as usual, i.e., by assigning a complex load value $\bar{S}_{kt}^i = [X_k]_t^i(1 + j \cdot tan\phi_k^i)$ to each load bus $i \in N\backslash$ {Slack Bus}, and finding the subsequent complex voltages $V_{kt}^i e^{j\delta_{kt}^i}$ for each i-th bus. Usual solution approaches rely upon the Newton-Raphson method [6] for meshed operating grids and upon the forward-backward sweep methods for radial operated grids [7].

Based on the power-flow solutions obtained, branch currents can be computed and be compared to grid equipment's capacity to identify congestion risks. A profile of currents in branch a-b, $[I_{kt}]^{ab}$ can be obtained by computing the current in each and every branch of the grid in each time period k and year t as:

$$[I_{kt}]^{ab} \cong (V_{kt}^a e^{j\delta_{kt}^a} - V_{kt}^b e^{j\delta_{kt}^b})(G^{ab} + jB^{ab}), k \geq 0; t = 1, \dots, H$$

Where, G^{ij} and B^{ij} are the real and imaginary parts of the element in the bus admittance matrix Y_{BUS} corresponding to the i-th row and j-th column.

The profile of currents is an indirect result of the sampling process. The flows result from the different dynamics of the different loads in the feeder. Being a result of the sampling, one may use the profile-flows to extract information about the distribution function of the current in that branch. Synthetized information about the distribution function can be obtained by frequency analysis in the domain of current-flows, I_{ky}, and be represented by a histogram. We give some details of histogram construction in the context of the presentation of results in the following.

Synthesis

As previously described, each DNS is characterized by twelve sampled load profiles with 96 time periods. As such, to simulate the operation in one year, we perform at least 1152 power flows, thus obtaining the values of currents and voltages for all branches and nodes in each of the 1152 time periods. To obtain such results over real-sized grids in very little time, one requires efficient, high-performance analysis algorithms; and to understand the results obtained, one requires specific frequency domain illustration capabilities.

DPlan has been extended in functionality to tackle those issues. It has evolved so as to report system-wide statistics for the whole grid and to provide specific risk measures for each and every branch and node of the system. System-wide statistics include:

- Yearly supplied energy (aggregated value from the weighted average of daily simulations);
- Yearly energy losses (also by aggregation);
- Expected energy not supplied (by aggregation);
- Capacity violation probability for grid assets;
- Maximum capacity violation in grid assets;
- Voltage violation probability given installation specific voltage limits;
- Maximum voltage violation in grid installations.

Specific risk measures were also reported for grid assets individually, both qualitatively and quantitatively. Qualitative results were reported in a simple, intuitive way over the geographic view of the grid by:

- Colouring the grid branches according to their capacity margin, i.e., the difference between each asset rating and the estimated maximum current for a given confidence level for such maximum (95%, in this case);
- Colouring the grid nodes and installation sites according to voltage margins, based on the estimated maximum and minimum voltages for a given confidence level.

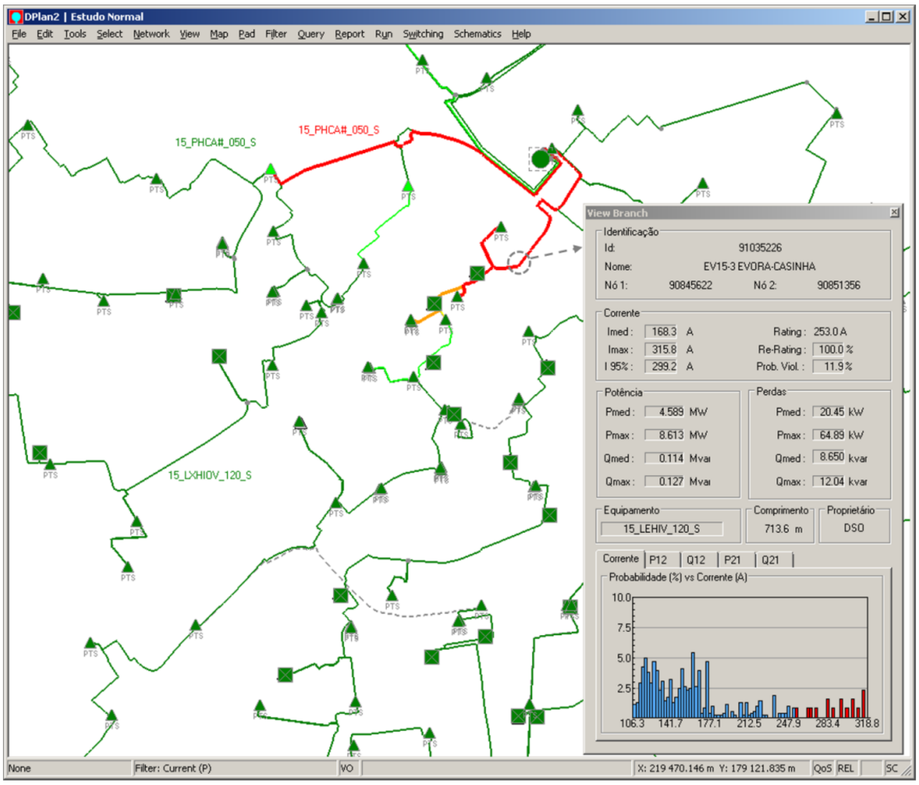
The colouring scheme has been made customized by risk level, in view of the reference values for risk established by the utility and regulators. In the case illustrated, colouring has been carried out based on the risk definition of Fig. 6.

Qualitative results are provided as frequency distribution of current and power for each branch and of voltage for each node/site. Once defined the range of possible values for a given branch or node result, say current $[I_{kt}]^{ab}$, the range is partitioned into a number of mutually disjoint intervals called buckets (or bins) and the frequency f_i is computed by counting values in each interval i. The set of pairs $\{(i, f_i)\}$ is the histogram $[I_{kt}]^{ab}$. In Fig. 5 we show a histogram of the currents in a particular branch (highlighted in white in the figure).

Quantitative results are provided as estimated maximum or minimum for a given variable (current, power, or voltage), as well as probabilities of violating ratings, regulatory voltage limits, etc. In Fig. 5, the currents turn out to exceed the branch rating with significant confidence – see that in the histogram. Such significance is quantified by estimating the probability of violating assets' capacity (11.9% in the figure) and is illustrated by colouring the histogram frequency bars for which the interval is beyond the branch rating (bin bars coloured in red).

Other quantitative results are presented and illustrated for the branch dialog of Fig. 5. Maximum and average current values estimated are reported for the branch, as well as the maximum and average power and maximum and average power losses.

Similar results are presented for estimated node and site voltages.



Identificação
Id: 91035226
Nome: EV15-3 EVORA-CASINHA
Nó 1: 90845622 Nó 2: 90851356

Corrente
Imed : 168.3 A Rating : 253.0 A
Imax : 315.8 A Re-Rating : 100.0 %
I 95% : 299.2 A Prob. Viol. : 11.9 %

Potência
Pmed : 4.589 MW
Pmax : 8.613 MW
Qmed : 0.114 Mvar
Qmax : 0.127 Mvar

Perdas
Pmed : 20.45 kW
Pmax : 64.89 kW
Qmed : 8.650 kvar
Qmax : 12.04 kvar

Equipamento: 15_LEHIV_120_S
Comprimento: 713.6 m
Proprietário: DSO

X: 219 470.146 m Y: 179 121.835 m

Fig. 5 – Illustration of the results obtained for the grid sampled loads after being synthesized. Results of the synthesis are shown in two different ways: (i) qualitative results are shown over the geographic view of the grid by colouring the grid conductor equipment according to the risk of capacity violations as defined in Fig. 6; (ii) quantitative results are shown for one selected branch -- one for which the current is expected to exceed the branch rating capacity with significant probability. Significance is illustrated by estimating the probability of capacity violation (11.9% in the figure) and by colouring the histogram frequency bars that corresponds to pairs for which the bin limits exceed the branch rating.

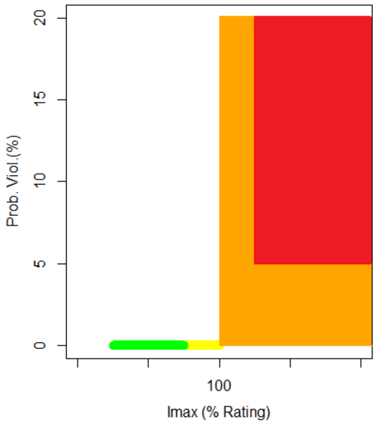
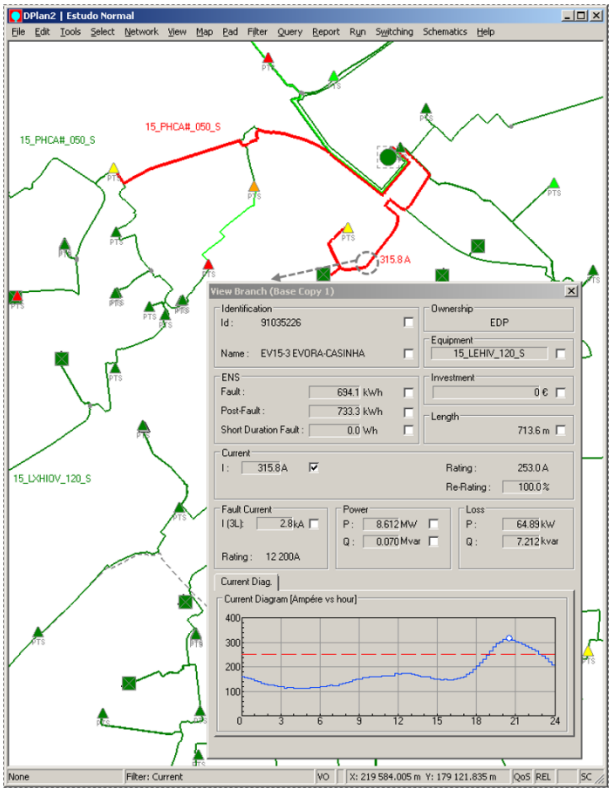


Fig. 6 – Risk definition for colouring branch rating filters in the probabilistic grid analysis. Rating violation magnitude (impact) is measured against violation chance (probability) to produce a risk index (colour).



Identification
Id : 91035226 □
Name : EV15-3 EVORA-CASINHA □

Ownership
EDP

Equipment
15_LEHIV_120_S □

ENS
Fault : 694.1 kWh □
Post-Fault : 733.3 kWh □
Short Duration Fault : 0.0 Wh □

Investment
0 € □

Length
713.6 m □

Current
I : 315.8A ☑
Rating : 253.0 A
Re-Rating : 100.0 %

Fault Current
I (3L): 2.8kA □
Rating : 12 200A

Power
P : 8.612 MW □
Q : 0.070 Mvar □

Loss
P : 64.89 kW
Q : 7.212 kvar

None | Filter: Current | VO | X: 219 584.005 m Y: 179 121.835 m | QoS | REL | SC

Fig. 7 – Illustration of chronologic power-flow analysis results for the branch selected in Fig. 5.

Synthetized results are crucial to pinpoint stressed grid operation conditions but are frequently insufficient to assess their severity level. Severity involves time-duration, and time has been forced to disappear in the frequency based synthetized mode of analysis. To understand the conditions in which the asset's capacity or voltage bounds are violated, one has to analyse the network in chronological mode. Navigation capabilities have been developed in DPlan to interchange between synthetized analysis and chronological analysis.

Fig. 7 illustrates the chronological underlying results obtained in a particular winter business day, before being synthetized in Fig. 5.

CONCLUSIONS

This paper addresses the strategy devised by EDPD to take advantage of the information contained in AMI data. The steps taken to cleanse, stratify and classify AMI data as typical load profiles were outlined. Specific data analytics tools were presented that characterize representative profiles and extract representative behaviours to be modelled as Markov chains.

Using Markov chains, the software applications that support decision-making in EDPD were evolved to simulate representative stochastic grid behaviour and provide probabilistic results with specified confidence levels on capacity use and voltage compliance.

Probability results are at the basis of risk controlled decision making, which **today** is crucial for a responsible network planning under the higher uncertainty associated with the deployment of renewables and distributed energy resources in distribution grids.

REFERENCES

[1] "Annual electric power industry report" (Nov. 2016). Tech. rep. U.S. Energy Information Administration. url: http://www.nowpublishers. com/ (accessed on 07/21/2014).

[2] V Pereira, P Mousinho, M L Jorge, "Identification of Electrical Energy Consumption Patterns", CIRED, Glasgow, Jun 2017.

[3] J A C Machado, P M S Carvalho, L A F M Ferreira, "Building Stochastic Non-Stationary Daily Load/Generation Profiles for Distribution Planning Studies" IEEE Trans. on Power Systems Vol. 33, No. 1, pp. 911-920, Jan. 2018.

[4] S. Asmussen, P. W. Glynn, Stochastic Simulation: Algorithms and Analysis. Series on Stochastic Modelling and Applied Probability, Vol. 57, Springer-Verlag New York, (2007).

[5] L A F M Ferreira, P M S Carvalho, C A Santos, J R da Silva, R Prata, F Carvalho "EDP's Decision Support Approach to Planning LV Smart Distribution Networks with DER," CIRED Workshop 2012, Lisbon, Portugal, May 2012.

[6] B. Stott, "Review of Load-flow Calculation Methods", IEEE Proceedings vol. 62, pp. 916-929, July 1974.

[7] A. Augugliaro, L. Dusonchet, S. Favuzza, M. Ippolito, E. R. Sanseverino, "A backward sweep method for power flow solution in distribution networks", Int. J. Elect. Power Energy Syst., vol. 32, no. 4, pp. 271-280, May 2010.