DEALING WITH SPARSE SMART METERING DATA IN TECHNO-ECONOMIC ANALYSIS OF LOW VOLTAGE NETWORKS

François VALLEE, Martin HUPEZ, Jean-François TOUBEAU, Zacharie DE GREVE
Electrical Power Engineering Unit, Faculté Polytechnique, University of Mons – Belgium
francois.vallee@umons.ac.be

ABSTRACT

Many Distribution System Operators (DSO) are currently working on the deployment of Smart Meters to ensure, among other things, a better observability of their system. However, in the context of Low Voltage (LV) networks, every customer does not benefit from such a measurement device, yet. Practically, when evaluating the techno-economic potential of an investment decision (cable upgrade, installation of storage units…), probabilistic load flows generally based on Pseudo-Sequential Monte Carlo algorithms can be used. Within this framework, the quality of the consumption/generation stochastic models has a major impact on the accuracy of the collected reliability indices. However, microscopic loads are often modelled with Synthetic Load Profiles that are rather representative of an aggregated consumption behaviour. The present paper aims at evaluating the benefits of a modelling strategy based on the definition of reference Cumulative Distribution Functions (CDFs) by applying it to a LV feeder in Belgium. Practically, this strategy requires a pre-processing clustering step. Moreover, a sensitivity analysis on the number (per cluster) of customers having a SM in their installation is complementarily conducted to evaluate the robustness of this modelling process in the context of techno-economic studies.

INTRODUCTION

The operation of Low Voltage (LV) networks is currently undergoing significant changes driven by the worldwide desire to support and facilitate the energy transition. Indeed, the strong desire to reduce greenhouse gases emissions has led to the emergence and integration of renewable energies into the system [1]. These distributed generators, predominantly photovoltaic (PV) panels at the LV level, are characterized by a strong random behaviour that results in high power fluctuations. Moreover, in addition to the electrification of the heating (i.e. heat pumps) and transportation (i.e. electric vehicles) sectors, new types of loads offering increased efficiency and controllability are modifying traditional end-users load profiles. The distribution networks were not originally sized for such operating conditions and the risk of technical issues such as voltage violations and congestions is clearly increasing. To ease the integration of this renewable energy based generation, suited technical solutions (storage, load shifting…) need to be deployed. Practically, to evaluate those solutions and to undertake adequate investment decisions, techno-economic analysis based on a probabilistic load flow framework (e.g. Monte Carlo simulation) has to be computed [2-3]. Those stochastic tools require the use of reliable models for every LV customer. In that way, the deployment of smart metering (SM) devices at the end-user level can be an opportunity. However, their actual roll-out is confronted to technical, financial as well as social barriers, and is therefore still limited to some sparse areas. Therefore, the authors [4] have developed a modelling strategy that firstly focuses on the segmentation of the end-users' consumption into representative clusters (components). Then, within each cluster, all available SM information is used to extrapolate statistical profiles thanks to an innovative load modelling methodology based on the definition of reference CDFs. The present paper aims at evaluating the benefits of this modelling strategy in the framework of a Pseudo-Sequential Monte Carlo simulation applied to a LV feeder in Belgium. To proceed, a benchmark is firstly defined. The latter is obtained by using, as inputs of the Monte Carlo algorithm, the historical measurements that are supposed to be available at all nodes of the tested feeder. Then, the improvement of the new modelling method compared to currently used approaches, such as Synthetic Load Profiles [5], can be highlighted. A sensitivity analysis on the number (per cluster) of customers having a SM in their installation is also conducted (progressive decrease of the number of SM-equipped customers within each defined cluster) and the obtained reliability indices (probability of overvoltage, mean value of the unbalance factor within the network…) are compared with the ones computed thanks to the benchmark. Consequently, potential limits (in terms of number of SM-equipped customers per cluster) to the applicability of the modelling strategy in the framework of a techno-economic analysis can be stated.

NEW LOAD MODELLING STRATEGY [4]

The proposed formulation is dedicated to a Pseudo-Sequential Monte Carlo framework. This approach is privileged as it demonstrates high accuracy for modelling hourly and daily seasonality while satisfactorily capturing the short-term dependence between successive time steps (a quarter of an hour in the present case). The global methodology can be decomposed into three complementary steps that are summarized in Figure 1:

-A preliminary analysis and classification of LV loads and generation units (mainly PV) into $n_c = n_{cD}$ *(load)* + n_{cG} (generation) clusters;

-The second step of the process is to statistically characterize the customers pertaining to the same cluster. This operation requires the definition of a reference CDF for each cluster. The latter is obtained thanks to data coming from consumers that have a Smart Meter and belong to the characterized cluster. Practically, n_c reference CDFs are thus built for each quarter of an hour of a monthly typical day (by doing this, the techno-

economical behaviour of the system has thus to be separately studied for each month of the year). Practically, since the clients pertaining to the same cluster can be of different sizes, the quarter-hourly SM measurements for each customer need to be standardized (with the yearly consumed energy E_{nd}^{tot} at the corresponding node n_d as this information is available whatever the type of meter) before the reference CDFs are established;

-Finally, the computed reference CDFs are included in a Pseudo-Sequential Monte Carlo tool, a given reference CDF being applied to the clients that pertain to the same cluster. Moreover, for a given quarter of an hour, the consumption behaviour of two consumers can logically be considered as totally independent (there is no reason justifying that two different customers behave similarly for each quarter of an hour of the day). Consequently, even if the same standardized reference CDF is used for all the customers belonging to the same cluster, the sampling on this reference CDF is made independently for each of those clients. The de-standardization step is then conducted and combined with a conversion into power of the sampled quarter-hourly consumed energies (with the assumption of constant consumption during each considered quarter of an hour).

Note that, for the present study, only the potential of the proposed modelling strategy from the consumption point of view is tested. The PV generation is modelled by use of historical measurements. This is justified by the fact that, in a geographically narrow LV distribution system and on a quarter hourly basis, PV generation units can be considered as strongly correlated [6]. It can therefore be expected that all the generation units are placed in a single cluster and that the proposed modelling strategy appears to be particularly robust from the generation point of view.

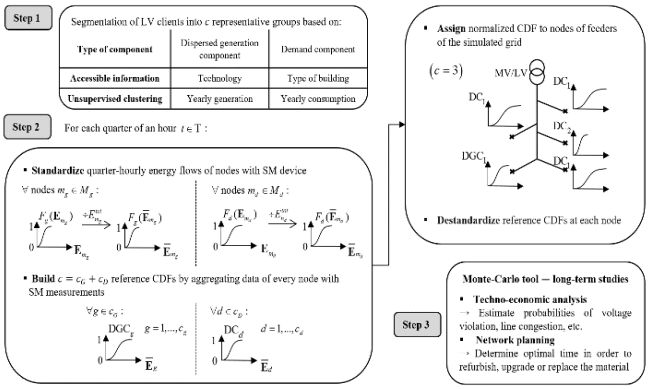


Fig. 1. General principle of load and generation modelling strategy

METHODOLOGY FOR THE SENSITIVITY ANALYSIS AND TESTED LV SYSTEM

Methodology for the sensitivity analysis

As already mentioned in the introduction, one first idea of this contribution is to evaluate the benefits of the load modelling strategy proposed in the previous section in comparison with traditional SLP models. A second objective is to evaluate its robustness when Smart Meters are not installed everywhere in the LV network but are, in some places, replaced by traditional meters (with a single annual energy index). To complete this analysis, the following steps are conducted:

-Step 1: All consumers connected to the LV system under study are supposed to be equipped with a Smart Meter and a benchmark is defined by conducting a Sequential Monte Carlo simulation that takes historical quarter-hourly measurements as direct inputs for the load models;

-Step 2: A pre-analysis for dividing the consumers based on their characteristics is carried out. Groups are firstly intuitively defined on basis of the type of clients: residential, farms, tertiary sector…. Concerning domestic customers, a subdivision using one-dimensional clustering based on their total annual consumption is also realized. A *k-means* algorithm is implemented to proceed to this clustering step. Practically, this type of algorithm consists in an unsupervised classification of data that aims at minimizing the variance within the *k* created clusters (where *k* is the desired number of clusters which is fixed by the experimenter);

-Step 3: The proposed load modelling strategy is then applied to the same consumers as in steps 1 and 2 (all of them still being equipped with Smart Meters) and a Pseudo-Sequential Monte Carlo simulation is computed. The collected indices (probability of overvoltage, mean unbalance factor value within the network…) are afterwards compared with the ones obtained at step 1 (benchmark). A comparison with indices simulated thanks to a SLP modelling process for LV loads is also conducted in order to better emphasize the benefits of the proposed load modelling process.

-Step 4: The number of customers having a Smart Meter is progressively decreased and the reference CDFs are thus defined with a limited number of clients compared to the case of step 3. For the customers without SM, their quarter-hourly consumption is defined by sampling on the reference CDF of the cluster in which they were classified during step 2, these sampled values being de-standardized by use of their annual index readings. Different strategies are applied to test the robustness of the load modelling methodology in front of the amount of LV clients having a Smart Meter. In a first approach, Smart Meters are progressively removed within one single cluster (the other clusters still being constituted of clients with SM). This study is separately conducted for each cluster to pinpoint the most sensitive ones and to draw limits of applicability for the proposed load modelling methodology. In a second phase, SM are simultaneously removed from different clusters. The robustness of the modelling methodology is again tested in that case and its sensitivity to the reduction of SM devices is compared with the previous approach.

Tested system

The tested system is part of a real LV network in Flobecq (Belgium). It is composed of one main feeder and two subfeeders. All the branches of the system are 3-phase overhead lines (lineic resistance $r = 0.32\ \Omega/km$, lineic self-reactance $x_{ii} = 0.243\ \Omega/km$, lineic mutual reactance $x_{ij} = 0.16\ \Omega/km$) and 20 nodes can be found on this network. The 18 prosumers of reference [4] are supposed to be connected to the tested system. All those connections are single phase ones and are randomly distributed over the available nodes (table 1).

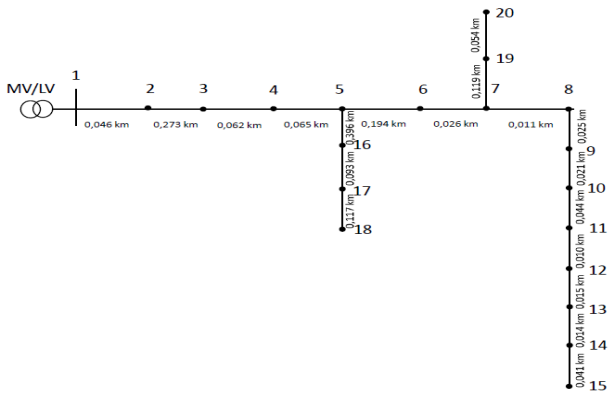


Fig. 2. Topology of the tested LV network

The pre-processing step is conducted identically as in reference [4]. The 18 loads are firstly categorized in domestic houses (13), farms (1) and shops (4). A *k-means* clustering algorithm (with $k = 3$ as this number of clusters was proved to be a good compromise between accuracy of the modelling process and size of the created clusters in reference [4]) is then applied to the domestic loads. Euclidean distances between the annual indexes associated to each of the 13 domestic customers are computed to define the different clusters. The clustering algorithm has grouped 5 customers into the first cluster which typically stands for traditional small families (the mean annual index for cluster 1 is equal to 3223.2 kWh/year). Then, the second group (the mean annual index for cluster 2 is equal to 5654.1 kWh/year) consists of 5 larger consumers (probably bigger houses with a higher number of inhabitants or with ageing appliances with lower efficiencies). Finally, cluster 3 (the mean annual index for this cluster is equal to 8432.7 kWh/year) includes clients that are probably equipped with electrical heating systems. To summarize, after the clustering step, $n_c = 6$ categories of LV loads are defined with $n_{cG} = 1$ PV generation class (composed of the 18 customers that are all equipped with PV panels) and $n_{cD} = 5$ consumption classes (4 shops, 1 farm, 5 small families (cluster 1), 5 bigger houses (cluster 2) and 3 customers probably equipped with electrical heating systems (cluster 3)).

Table 1. Characterization and connexion scheme of the different loads

Node	*Type of load*	*Connexion phase*	*Installed PV capacity (kVA)*
1	-	-	-
2	Cluster 2	C	3.5
3	Cluster 3	B	4.1
4	Cluster 1	B	2.63
5	Shop	C	3.4
6	Cluster 2	A	2.65
7	Cluster 1	A	2.65
8	-	-	-
9	Cluster 1	C	3.47
10	Shop	B	3.47
11	Cluster 3	A	4
12	Shop	C	2.65
13	Cluster 2	B	4.6
14	Cluster 1	B	4.2
15	Cluster 3	B	5
16	Cluster 1	C	5
17	Farm	B	5
18	Cluster 2	B	5
19	Shop	A	5
20	Cluster 2	A	2.65

Note that, as already mentioned, the modelling strategy will only be applied to consumption. Consequently, historical measurements will be considered for PV generation for all the tested cases.

SIMULATION RESULTS

Performance of the modelling strategy in front of traditional SLP approaches

In this section, when the new modelling strategy is applied, all the 18 clients are supposed to be equipped with SM. The collected indices after application of the Pseudo-Sequential Monte Carlo simulation are compared with the ones obtained with a direct utilization of the historical measurements (benchmark) and with the ones computed when SLP curves are applied to all loads. Table 2 summarizes the obtained results in terms of probability of (under- or over-) voltage violation and of unbalance in the entire network. Note that, to compute sufficiently high indices, voltage limits have been set to 220 V (lower limit) and 240 V (upper limit) for the computation of probabilities of voltage violations while the mean unbalance factor value is computed to estimate the unbalance of the network. Finally, the new modelling strategy is applied to characterize the load behaviour during the month of August (definition of a typical day of August). Consequently, historical measurements for August (2013 until 2015) and a SLP profile for the same month [7] are considered to respectively compute the reliability indices of the benchmark and of the SLP approach. It can be observed in table 2 that the new modelling strategy outperforms the SLP one. This result

can be easily explained as the SLP method, by associating the same mean consuming behaviour for all the customers connected to the tested distribution system, leads to worsened indices. Indeed, by applying the SLP approach, the mutual influence of all the residential customers, whatever their location in Belgium (urban, rural…), connected at the distribution level impacts our tested system and logically involves worsened reliability indices.

Table 2. Simulated reliability indices (for the entire network) with the benchmark, the new modelling methodology and the traditional SLP approach for the month of August (48th quarter of an hour of the day)

	Probability of voltage violation [%]	*Mean unbalance factor value [%]*
Benchmark	4.21 (a) / 1.75 (b) / 0 (c)	0.27 (a) / 0.60 (b) / 0.86 (c)
SLP approach	20.85 (a) / 28.25 (b) / 0 (c)	0.22 (a) / 0.95 (b) / 1.09 (c)
New modelling methodology	3.33 (a) / 1.82 (b) / 0 (c)	0.35 (a) / 0.74 (b) / 1.06 (c)

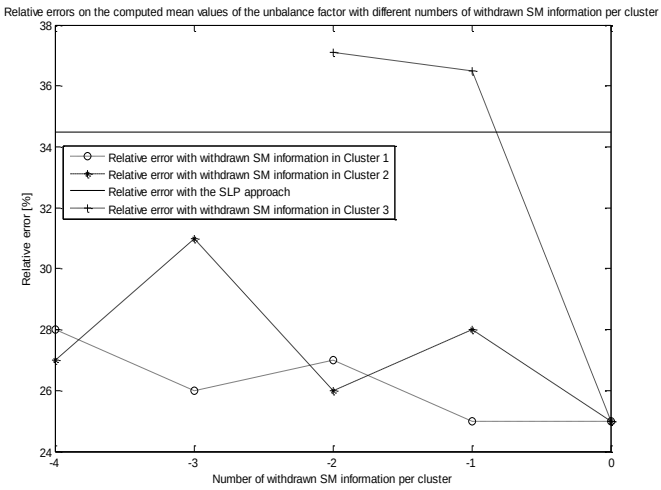
Table 3. Simulated reliability indices for node 18 (same working conditions as in table 2)

	Probability of voltage violation [%]	*Unbalance factor value [%]*
Benchmark	0 (a) / 20 (b) / 0 (c)	0.02 (a) / 1.22 (b) / 1.18 (c)
SLP approach	3.33 (a) / 40 (b) / 0 (c)	0.15 (a) /1.34 (b) / 1.19 (c)
New modelling methodology	0 (a) / 19.51 (b) / 0 (c)	0.01 (a) / 1.28 (b) / 1.26 (c)

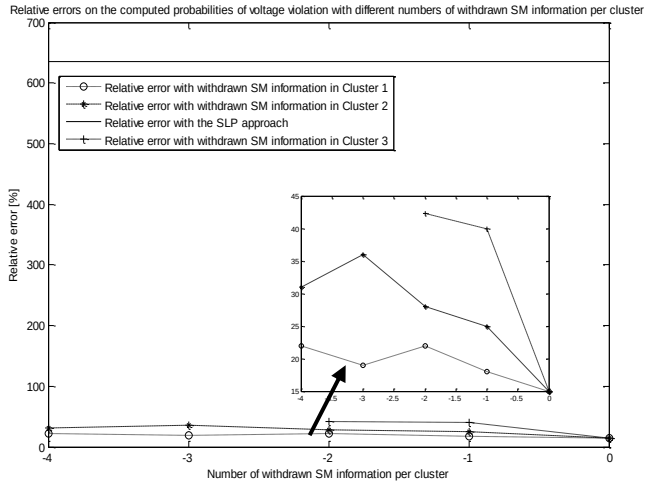
Table 3 particularises the collected indices for node 18 which is one of the most critical nodes of the tested system (the installed PV generation for this client is slightly oversized compared to its consumption and, moreover, it is connected at the end of a subfeeder). For this node and, for the same reasons as the ones pushed forward during the global analysis of table 2, the same observations can be made when comparing the indices computed at node 18 with the SLP approach and with the new modelling strategy.

Performance of the modelling strategy with sparse smart metering data

In a first step, it is supposed that smart metering information is progressively removed within one single domestic cluster (while the others keep all the information to build the reference CDFs). The same operation is repeated for each domestic cluster and the relative errors (in absolute value) on the collected indices (by comparing them to the ones obtained with the benchmark in table 2) are provided in Figure 3. It can be observed that, when dealing with domestic loads, the proposed modelling approach is quite robust to the number of clients that are equipped with a SM. Indeed, a limited degradation in the collected indices is observed in Figure 3 when SM information is progressively removed within each of the three predefined residential clusters; only cluster 3 (composed of 'big' consumers) leading to a larger (but acceptable as it globally stays below the order of magnitude of the relative errors computed with the traditional SLP approach) volatility in the collected indices. This result seems to demonstrate that the quarter-hourly load distributions of domestic customers are very similar and, therefore, very close in terms of distance. Consequently, reference CDFs are only slightly impacted when SM information is withdrawn in their elaboration process. Note that the same conclusions can be drawn when performing such an analysis on critical node 18.



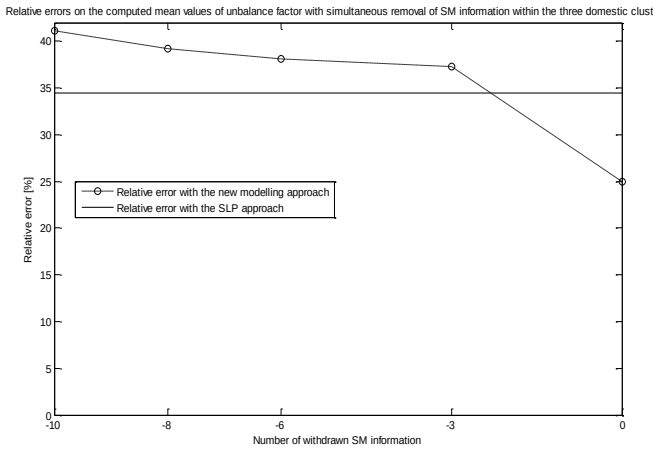
(a)



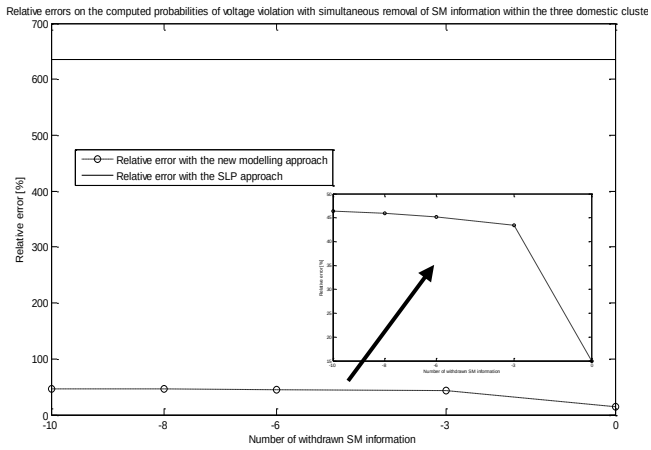
(b)

Fig. 3. Mean global (on the three phases) relative errors on the simulated reliability indices with the traditional SLP approach and the new modelling strategy (with SM information progressively removed within one domestic cluster) for the month of August (48th quarter of an hour of the day): a) unbalance factor / b) probability of voltage violation

In a second and final analysis, smart metering information is progressively removed from all the domestic clusters at the same time (starting thus by 3 removed clients, one in each cluster). The collected relative errors are provided in Figure 4. The same observations as in Figure 3 can be made and the greatest influence (on the collected reliability indices) of the clients pertaining to cluster 3 is confirmed. Indeed, after 2 out of the 3 clients pertaining to cluster 3 have been removed (-6 clients in the overall), relative errors do not evolve significantly anymore.



(a)



(b)

Fig. 4. Mean global (on the three phases) relative errors on the simulated reliability indices with the traditional SLP approach and the new modelling strategy (with SM information progressively removed from the three domestic clusters at the same time) for the month of August (48th quarter of an hour of the day): a) unbalance factor / b) probability of voltage violation

CONCLUSION

In this paper, the performances of a new load modelling strategy relying on SM data have been evaluated in the framework of a Pseudo-Sequential Monte Carlo techno-economic analysis conducted on a LV feeder. Those performances have been compared not only to the ones obtained with traditional SLP models but also to the ones reached with the new modelling methodology when SM devices are not installed within each domestic customer. By doing this analysis, the new load modelling strategy proved to be more accurate than the SLP approach by being able to better approximate the local behaviour of LV clients. Moreover, thanks to a pre-processing clustering step that allows to categorize LV clients, this new modelling method also demonstrated its robustness in front of SM missing information and, therefore, its suitability for LV networks not (yet) fully equipped with SM devices.

REFERENCES

[1] T.J. Hammons, 2008, "Integrating renewable energy sources into European grids", *International Journal of Electrical Power & Energy Systems*. vol. 30, 462-475.

[2] F.J. Ruiz Rodriguez, J.C. Hernandez, F. Jurado, 2013, "Technical impact of photovoltaic distributed generation on radial distribution systems: stochastic simulations for a feeder in Spain", *Electrical Power Systems Research*. vol. 50, 25-32.

[3] S. Conti, S. Raiti, 2007, "Probabilistic load flow using Monte Carlo techniques for distribution networks with photovoltaic generators", *Solar Energy*. vol.81, 1473–1481.

[4] J-F. Toubeau, M. Hupez, V. Klonari, Z. De Grève, F. Vallée, "Statistical Load and Generation Modelling for Long Term Studies of Low Voltage Networks in Presence of Sparse Smart Metering Data", *42nd Annual Conference of Industrial Electronics Society*, IEEE IECON2016, Florence, Italy, 23-27 October 2016.

[5] P.F. Albrecht, M.P. Bhavaraju, B.E. Biggerstaff, R. Billinton, G.E. Jorgensen, N.D. Reppen, P.B. Shortley, 1979, "IEEE reliability test system", *IEEE Transactions on Power Apparatus & Systems*. vol. 98, 2047-2054.

[6] V. Klonari, J-F. Toubeau, J. Lobry, F. Vallée, 2016, "Estimating the photovoltaic hosting capacity of a Low Voltage feeder using Smart Meters measurements", *Smart Metering Technology & Services, Intech book*. chap. 7, 131-158.

[7] http://www.synergrid.be/index : link to download SLP profiles for Belgium.