Net Load Disaggregation and Forecast at Medium Voltage Level

Mateo Toro Cárdenas
mateo.toro@tecnico.ulisboa.pt
Instituto Superior Técnico, Universidade de Lisboa, Portugal
December 2021

***Abstract*—With more and more micro-generation being connected to the lower voltage levels, it becomes increasingly difficult for operators to anticipate the loading condition of their networks. Several facts concur to such difficulty, one being the frequent lack of knowledge about the generation component of the measured net load. The generation component is typically more volatile than natural load and is impossible to predict without reliable information on the grid renewable installed capacity. In this context, the objective of this thesis is to propose a non-intrusive methodology to estimate distributed generation' installed capacity. Relying upon net load historical data obtained at the secondary substation level and on historical meteorological data, this thesis presents algorithms that proved effective in determining an approximation of the renewable embedded capacity. The accuracy and effectiveness of the algorithms presented are illustrated with real data, for real use cases in distribution grids.**

***Index Terms*—Distribution Network, Forecast, Net Load Disaggregation, Photovoltaic generation, Secondary Substation**

NOMENCLATURE

d	Pair of days from consecutive weeks (same weekday)
D	All possible pair of days d in a data set
Δ	Precision in Modified Bisection
ε_d	RMS error between the estimate natural load $\hat{L}$ profiles of a pair of days d
$\varepsilon_{d,i}$	RMS error between the estimate natural load $\hat{L}$ profiles of a pair of days d per index i
$freq$	Number of times that a value is repeated in a data set
G	Maximum number of iterations in Modified Bisection
I	Solar irradiance
K	Scale Factor of $\hat{PV}$ size
L	True Natural Load
$\hat{L}$	Estimated Natural Load
$\hat{L}^{pred}$	Estimated Natural Load prediction
m	Statistical mode Mo of estimated DG capacity $\hat{x}_d$ in all possible pair of days D
Mo	Statistical Mode
N	True Net Load
N^{pred}	Net Load prediction
$\hat{N}^{pred}$	Aggregated Net Load prediction
$\hat{N}^{pred}_{partial}$	Partial Aggregated Net Load prediction
PV	True PV Generation
$\hat{PV}$	Estimated PV Generation
$\hat{PV}^{pred}$	Estimated PV Generation prediction
$x_{d,i}$	Installed DG capacity per pair of day d and index i
$\hat{x}_d$	Estimated installed DG capacity per pair of day d
X	True installed DG capacity
$\hat{X}$	Estimated installed DG capacity

I. INTRODUCTION

Distributed generation (DG) penetration is increasing mainly due to the deployment of wind and solar photovoltaic (PV) technologies. Most wind generation is being installed at the medium and high-voltage networks while PV is being installed across all voltage levels, including in low-voltage (LV) networks [1]. One of the main operational challenges of the deployment of PV at LV networks is the change in the consumption profile it causes at the higher voltage levels (duck curve) [2], which needs to be anticipated by the Transmission System Operators (TSOs) and mainly by Distribution System Operators (DSOs) to avoid voltage constraints (under and over voltages) in the system.

Anticipating such profile changes with acceptable accuracy requires knowledge of the approximated PV installed capacity in order to infer about PV power active power generation and its dependencies on meteorological conditions. However, such capacity is frequently unknown to the DSO, either because it did not keep track on the licensed PV connections, or because such connections did not require any kind of licensing from the DSO, which is frequently the case for small capacities and/or self-consumption in many EU countries. In [3], an overview of the current requirements for grid connections in several European countries is presented; in [4], the interconnection procedures of small distributed generation in USA are described.

In this thesis, an effective non-intrusive methodology that is capable of estimating the PV capacity installed in a LV distribution grid is presented. The methodology makes use of the secondary substation historical data of active power measurements (net load) and available weather data to infer about the approximated installed capacity by disaggregating load and PV generation active power profiles. Several works have addressed challenges related to non-intrusive load disaggregation (NILM). For example, in [5] a disaggregation method is applied to residential buildings, including heating ventilation and air conditioning systems that makes use of

thermodynamic models. Other publications refer to the benefit of providing information about feeder's load and network demand to the grid operator by disaggregating losses and reactive power injections separating feeder demand into its components [6]. Different approaches and procedures are use to undertake disagreggation. Examples are online learning algorithms, Dynamic Fixed Share, and linear regression models with complex objective functions [7].

Regarding NILM, literature is extensive. From the first time it was presented in [8] its goal was to "determine the energy consumption of individual appliances turning on and off in an electric load". From that point, authors have made valuable contributions by providing detailed description of operational states of consumer appliances as described in [9]. However, even the most significant approaches involving Hidden Morkov Models (HMM) to describe loads and segmented integer quadratic constraint programming (SIQCP) to effectively solve the NILM problem are only focused on household power profiles [10]. On top of that, the applications for NILM are not fully described and addressed given various aspects such as constant market introduction of new appliances, reduced open common data bases, development and prototyping costs, scalability and lack of innovative business models [11].

A comprehensive review of NILM methods in the home energy management system is presented in [12] and the critical issues of NILM are discussed and the Advanced NILM (ANILM) concept is proposed, including the most important characteristics of this new method. Another review paper, more centred on the performance evaluation of NILM methods is presented in [13] aiming the comparison of the performance of different NILM methods applied to several datasets. The performance evaluation metrics for NILM methods are also reviewed in [13]. The problem of Datasets used in NILM methods is also addressed in [14]. According to the authors, after a critical analysis of 42 NILM datasets, the performance and use of NILM methods are limited by the features that are available in each dataset as well as by the lack of standards used to describe these features.

The result is a handful of questions and topics that are left open to be validated in the future. Specifically, this thesis aligns with the opportunity, based on all the previous knowledge, to escalate the applications of the disaggregation concept merged with the increasing use of DG involving renewable energy sources (RES) and the option to provide a valuable final use for DSOs involving an improved methodology to forecast load.

The present thesis is divided into five sections. The core of the development is presented on Section II, describing the methodology used. Illustration of the proposed developments is presented on Section III. Section IV presents application results for a real case, showing how the methodology can effectively be used to improve load forecasting. Finally, conclusions are presented in section V.

II. Methodology

This thesis proposes a new methodology allowing the active power consumption and generation disaggregation at the secondary substation (SS) level. The method uses two inputs. First, Net Load, N, i.e. active power measurements which could be in the downstream and/or upstream direction, from a meter installed in SS. Second, weather information, specifically, solar irradiance, I. The method produces two outputs. First, "Estimated Installed DG capacity", $\hat{X}$, an estimation of the capacity of the micro-generation installed in the low voltage network and connected at the SS. Second, "Disaggregated Profiles", the dissaggregated profiles of the N's components, i.e. Estimated Natural Load, $\hat{L}$, and Estimated PV generation, $\hat{PV}$. The first considers the power demand of consumers connected downstream from the meter, while $\hat{PV}$ refers to the power delivered in the upstream direction (reverse power flow), by the DG installed under the meter. As seen in Fig. 1, the input data is processed by two different algorithms, "Day Comparison" and "DG Size Estimation", which will be explained in detail in Section II-A and II-B, respectively. The outputs of this methodology can be used for different purposes. In the present thesis, the Disaggregated Profiles, i.e. $\hat{L}$ and $\hat{PV}$ profiles, will be used as inputs for load forecasting method.

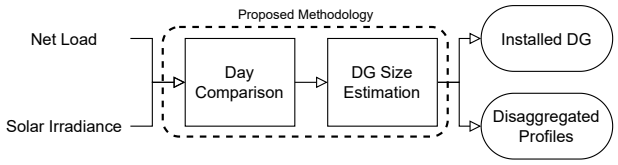


Fig. 1. Net-Load dissagregation methodology

A. Day Comparison

The name of this algorithm comes from the action of taking two days of data to compare them. The days are taken in pairs of the same weekday of consecutive weeks, i.e. Monday Week 1 with Monday Week 2, Tuesday Week 1 with Tuesday Week 2, and so on, then Monday Week 2 with Monday Week 3, and so forth sweeping the whole data set. The pair of days are referred as d and there will be around 173 possible pair of days in a six months, or 180 days, data set. Each comparison produces an estimated installed DG capacity per pair of days analyzed, $\hat{x}_d$.

The comparison has the objective of finding the value $\hat{x}_d$ that minimizes the Root Mean Square Error (RMSE), ε_d, between the natural load profiles $\hat{L}$ of the pair of days analyzed. Load profile $\hat{L}$ is obtained based on the estimation of PV profile $\hat{PV}$ by using equation (2) and the net-load profile N. The estimation of the PV profile $\hat{PV}$ obtained based in the irradiation measurements I by the equation (1). After finding $\hat{L}$ profiles of both days, ε_d is calculated.

$$\hat{PV} = \hat{x}_d I \tag{1}$$

$$\hat{L} = N + \hat{PV} \tag{2}$$

The minimization problem is solved with a modified bisection method presented in Algorithm 1. However, the modified bisection goal here is not to find a function's root [15], but rather to find the function's minimum. As usual in such method, three initial points for $\hat{x}_d$ are taken. In this case, lower, upper and midpoint guesses are $x_{d,0}$, $x_{d,1}$ and $x_{d,2}$, respectively. The function f in Step 1 refers to the above explained comparison carried out for each of the DG guesses, i.e., take $x_{d,i}$ and obtain $\varepsilon_{d,i}$ where $i = 0, 1, 2$.

Algorithm 1 Modified bisection in Day Comparison

Input: Lower and Upper DG size guesses, $x_{d,0}$ and $x_{d,1}$.
Output: $\hat{x}_d$ and ε_d.

1: Define function $f(x_{d,i}) = \varepsilon_{d,i}$; $\forall i = 0, 1, 2$
2: $g = 0$
3: **while** $g < G$ or $\delta > \Delta$ **do**
4: $x_{d,2} = (x_{d,0} + x_{d,1})/2$
5: $\varepsilon_{d,i} = f(x_{d,i})$; $\forall i = 0, 1, 2$
6: **if** $[\varepsilon_{d,0} < \varepsilon_{d,1}]$ and $[\varepsilon_{d,2} < \varepsilon_{d,1}]$ **then**
7: $x_{d,1} := x_{d,2}$
8: **else**
9: $x_{d,0} := x_{d,2}$
10: **end if**
11: $\delta = abs(x_{d,1} - x_{d,0})$
12: $g = g + 1$
13: **end while**
14: $\hat{x}_d = x_{d,i} : min(\varepsilon_{d,i})$
15: $\varepsilon_d = \varepsilon_{d,i} : min(\varepsilon_{d,i})$
16: **return** $\hat{x}_d, \varepsilon_d$

There are two stopping criteria: number of iterations, g, and precision, δ. When one of those is reached, the outputs of the modified bisection are $\hat{x}_d$, the point $x_{d,i}$ with the least $\varepsilon_{d,i}$ where $i = 0, 1, 2$, and ε_d, the least $\varepsilon_{d,i}$ as well. Then, the following pair of days, d, is analyzed until the end of the dataset is reached.

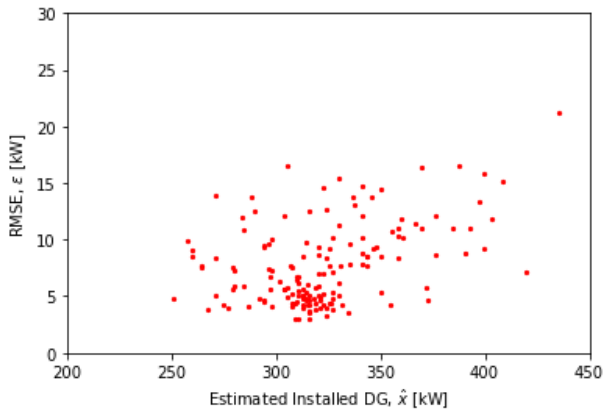


Fig. 2. Day comparison output

Day Comparison algorithm's output is the collection of $\hat{x}_d$ and ε_d for every d – see Fig. 2. The length, D, of that set is the original number of days subtracted of 7, as the analysis is carried out for pairs of days of same weekday in consecutive weeks.

B. DG Size Estimation

The second algorithm groups the points obtained in the Day Comparison algorithm identifying a single one as the best representative of the whole set. The output of the algorithm is the "Estimated Installed DG", $\hat{X}$. Such identification process is done according to the origin of the data used in the methodology. Specifically, the PV data in this thesis can be considered synthetic or real as will be described in detail in section III. Additionally, the input of DG Size Estimation is the collection of points $\hat{x}_d$ and ε_d for $d = 0, .., D$.

When the PV generation data is considered synthetic, a Mode/Cluster algorithm is executed. Such can be seen in Algorithm 2 and was developed by the author of this thesis. Its execution starts by finding the statistical mode, Mo, of the input data. If the frequency, $freq$, of such mode contents more than 10% of the points, the mode is given as DG Size Estimation's answer. Otherwise, an iterative histogram is executed.

Algorithm 2 Mode/Cluster Algorithm

Input: $\hat{x}_d$, $\forall d = 0, .., D$
Output: $\hat{X}$

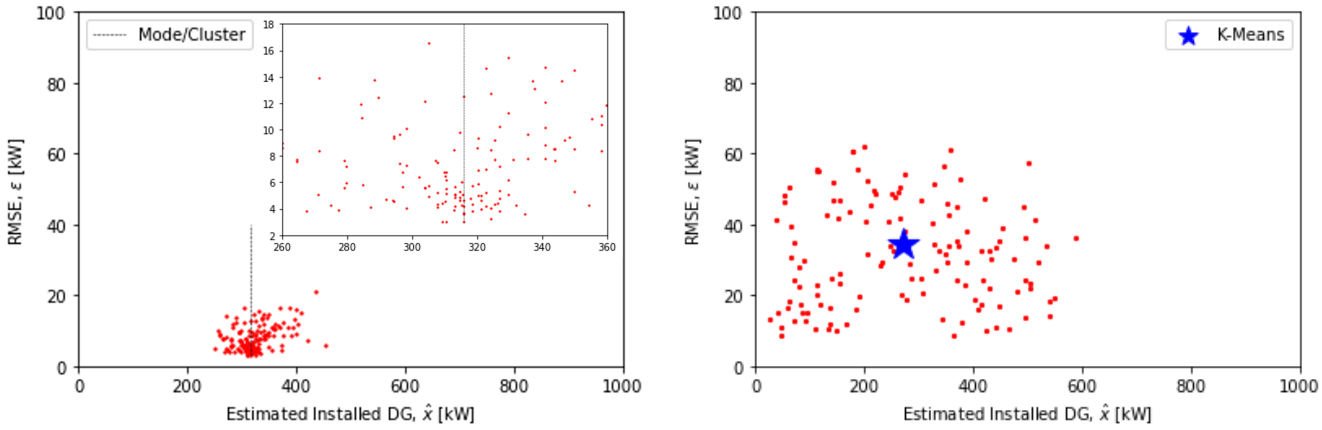
1: $m = Mo(\hat{x}_d \forall d = 0, .., D)$
2: **if** $freq(m) \geq 10\%$ **then**
3: **return** $\hat{X} = m$
4: **else**
5: $\hat{X}$ = Iterative Histogram Answer
6: **end if**
7: **return** $\hat{X}$

An iterative histogram is the process of taking the interval with most data repetitively until the distance between the extreme values of such interval are less than a given threshold stopping the iterations. In other words, in an recursive way make a histogram of the most representative interval of an initial histogram. The midpoint of the resulting interval is consider the Estimated Installed DG, $\hat{X}$. The answer provided by Mode/Cluster algorithm is seen graphically in Fig. 3a. Synthetic PV uses results in a very grouped data set at the end of Day Comparison algorithm.

On the other hand, when the PV generation data is obtained in real PV plants, a K-means can present better performance. Its goal is to find a point that minimizes the distance among all the points being analyzed [16], [17]. The previous given the larger dispersion in the points at the end of Day Comparison algorithm as seen in Fig. 3b when using real PV.

III. CASE STUDY

The data used in this thesis is composed of 15 minute granularity measurements of various existing SS Transformers in the center region of Portugal ranging from 250 kW to 630 kW. Those measurements are made of downstream active



(a) (b)

Fig. 3. DG Size Estimation Output considering Synthetic (a) and Real (b) PV Generation

power. The weather variables are obtained from [18] located in the city of Faro, also in Portugal.

There are two scenarios considered in this thesis regarding the relation between solar irradiance and upstream PV generation. In the first case, a synthetic PV generation was created given a linear relation obtained by multiplying the solar irradiance by a factor K according to the desired DG installed capacity, X, to be analyzed [19]. On the other hand, the second scenario considered a real PV generation data of a panel with 220 Wp and the relation between the irradiance and the PV generation is not exactly linear. The previous was achieved by using real PV generation profiles provided by [20]. The PV generation can also be adjusted to the desired DG installed capacity X.

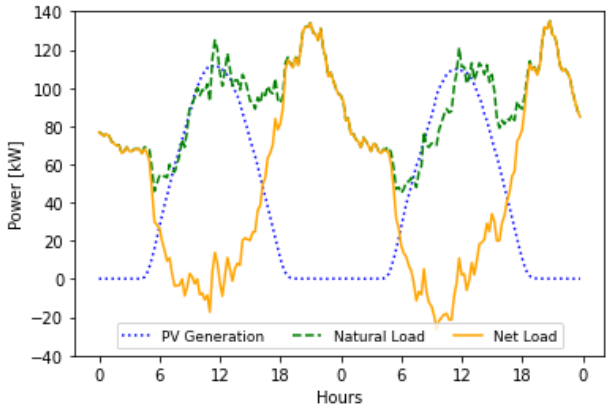


Fig. 4. Composition of load profiles in the case study

Having an user-defined PV generation size, X, a true PV generation, PV, profile can be created in the upstream direction, either synthetic or real shaped, by using I. The next step was to combine the PV profile with the original measurements of the existing downstream SS Transformers, i.e. True Natural Load L, in order to create N profile. The previous can be seen in (3) and (4).

$$PV = IK \tag{3}$$

$$N = L - PV \tag{4}$$

It is worth saying that the resulting N profile is not taken from a measuring device, it is constructed for the purpose of this thesis. During the disaggregation analysis, the goal is to obtain the $\hat{PV}$ and $\hat{L}$ profiles given the N profile as input. The described load composition is presented in Fig. 4. There, the positive values of Net Load means that the active power flows to downstream (from the medium voltage (MV) to the LV network) and the negative values means that the active power flows to upstream from (LV to MV network).

A. Synthetic PV

As explained previously, synthetic PV, in this thesis, refers to a linear relation between Solar Irradiance and PV active power generation, i.e. taking I data and multiply it by a factor K in order to create PV data. In such case, PV has the same shape as I but scaled by a factor K to reach the desired user-defined size X. After that, the N profile is created as previously described.

The methodology was tested in 10 SS by applying 5 different PV sizes corresponding to 10%, 25%, 50%, 75% and 100% of the SS Transformers' capacity. Table I shows the best, worst and average results, obtained by the proposed methodology, for the installed PV capacity estimation.

The algorithm proved to be effective and highly accurate in the tests it was subjected by applying different load profiles magnitudes and DG installed capacities, X. In average, the estimated PV size, $\hat{X}$, deviates ± 5.65 kW from the true value, X. This represents a 98.82% accuracy in the DG installed capacity estimation.

B. Real PV

In this scenario, real PV power generation measurements were taken into account, in other words, the relation between I

TABLE I
ESTIMATED INSTALLED DG CAPACITY RESULTS AT SS LEVEL UNDER SYNTHETIC PV SCENARIO

% of capacity	Best [kW] True (X)	Best [kW] Estimated ($\hat{X}$)	Worst [kW] True (X)	Worst [kW] Estimated ($\hat{X}$)	Average error [%]
10	63.00	65.52	25.00	30.03	1.29
25	157.50	162.44	62.50	66.89	0.75
50	315.00	309.87	125.00	129.68	0.92
75	472.50	473.68	187.50	191.11	1.07
100	630.00	635.38	250.00	248.44	1.85

and PV is not linear. In this context, a component of delay and operational uncertainties was considered in the calculations. The previous can be seen in the dispersion graph presented in Fig. 5, where the behavior of PV generation during one month is not perfectly linear in relation with the solar irradiance as proposed in the previous subsection. In addition, Fig. 6 shows, in more detail, how a real PV active power generation measurements does not match exactly the solar irradiance's profile shape.

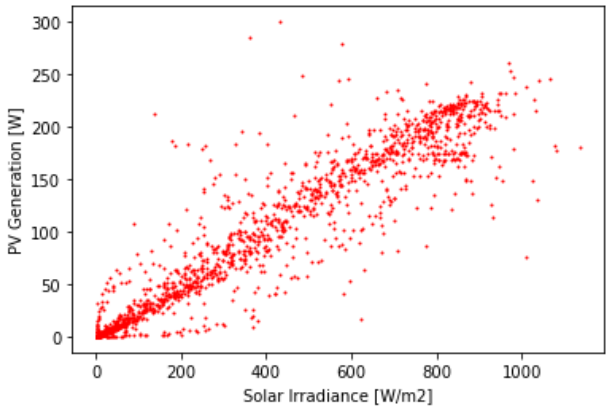


Fig. 5. Real PV generation on a 220 Wp panel against solar irradiance during one month

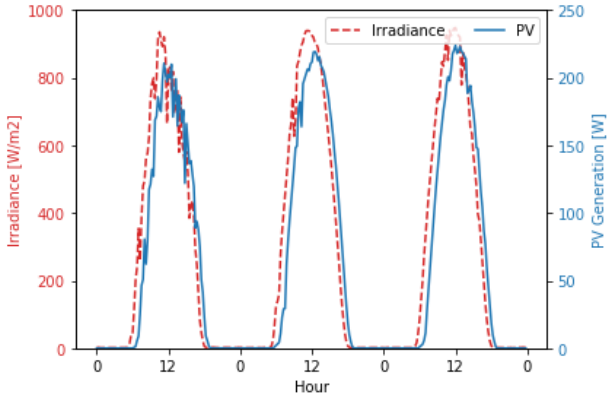


Fig. 6. Solar Irradiance and Real PV generation on a 220 Wp panel during 3 days

The original measurements given by [20] and seen in Fig. 5 correspond to a 220 Wp panel located in the city of Faro, Portugal. During the construction of this case study, the data was scaled linearly to the desired DG installed capacity to be analyzed.

Again, the methodology was tested in the same 10 feeders by applying five different PV sizes corresponding to 10%, 25%, 50%, 75% and 100% of the SS Transformers' capacity. The results can be seen in Table II. As expected, the methodology's performance decreased in accuracy. In average, the estimated PV size, $\hat{X}$, deviates ± 67.44 kW from the actual value X. This represents a 85.86% accuracy in the DG installed capacity estimation. Even though the effectiveness of the methodology in this second study was reduced, it can be considered accurate given the real world imperfections.

TABLE II
ESTIMATED INSTALLED DG CAPACITY RESULTS AT SS LEVEL UNDER REAL PV SCENARIO

% of capacity	Best [kW] True (X)	Best [kW] Estimated ($\hat{X}$)	Worst [kW] True (X)	Worst [kW] Estimated ($\hat{X}$)	Average error [%]
10	63.00	83.84	25.00	61.35	7.67
25	157.50	195.05	62.50	95.73	9.65
50	315.00	373.46	125.00	168.45	13.77
75	472.50	559.85	187.50	242.51	17.87
100	630.00	742.77	250.00	316.67	21.67

Fig. 7 shows the estimation' errors in PV size according to the percentage of penetration in the corresponding SS Transformer. The PV size does not have effect when using synthetic PV information. In that case, the methodology is very stable showing high precision. However, it does affect the methodology when dealing with real PV data. The bigger the DG installed capacity the more average percent error is expected.

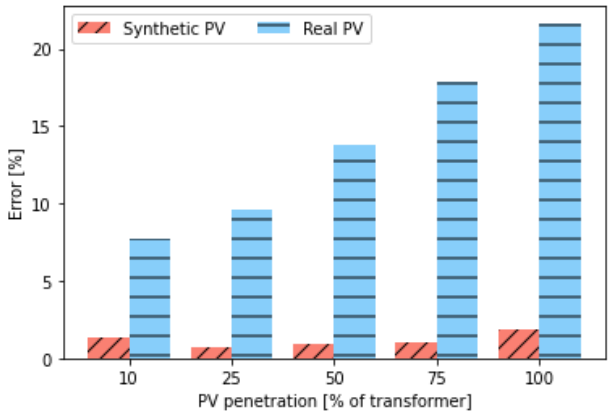


Fig. 7. PV Size Estimation Average Errors according to PV penetration

IV. APPLICATION TO LOAD FORECAST

The main goal of this section, is to evaluate if the the proposed methodology led to an improvement in both the consumption and generation forecasts, when compared to the performance obtained with the same methods while using the original Net Load measurements. The results obtained

with the disaggregation process were used as input for the consumption and generation forecast methods. Forecast were carried out using a classical Auto-regressive model with a single exogenous meteorological variable, i.e. an ARX model – see (5) for a possible expression of the autoregressive equation, [21]. In our case, the exogenous external variable used was the solar irradiance I.

$$Y_t = \beta_0 + \sum_{j=1}^{p} \beta_j Y_{t-j} + \sum_{f=0}^{s} \theta_f I_{t-f} + \xi_t \tag{5}$$

$where$:

p	Model order of Endogenous Variable
s	Model order of Exogenous Variable
Y_t	Value to predict for instant t
Y_{t-j}	j^{th} previous value of the variable to predict
I_{t-f}	f^{th} value of exogenous variable
β_j	Coefficient of the j^{th} endogenous variable lag
θ_f	Coefficient of the f^{th} exogenous variable lag
ξ_t	Stochastic term

In ARX models, the variable to predict, Y, is obtained as a linear combination, not only of the previous values of Y, but also of the history of external variable, I, outcomes, this way capturing the autocorrelation of Y together with its dependence on I. Both coefficients, the autoregression, β and the new coefficients for the exogenous variable, θ, are obtained with ordinary least-squares method [22], and are used to forecast future values of Y given future estimates for I.

TABLE III
DATES TO EXECUTE THE ARX MODEL

Training dates		Forecasting dates	
21/01/2017	6/06/2017	7/06/2017	13/06/2017

The ARX model was executed according to the dates in Table III. It can be seen that the model was trained with around six months of 15-minute data and was used to forecast seven days ahead with 15-minute resolution as well. In total, three variables were forecast: N, $\hat{PV}$ and $\hat{L}$ and the results were: N^{pred}, $\hat{PV}^{pred}$ and $\hat{L}^{pred}$, respectively. Afterwards, the $\hat{PV}^{pred}$ and $\hat{L}^{pred}$ were combined to create aggregated Net Load prediction, $\hat{N}^{pred}$ according to (6).

$$\hat{N}^{pred} = \hat{L}^{pred} - \hat{PV}^{pred} \tag{6}$$

At this point, there are two cases for forecasting, i.e. $\hat{N}^{pred}$ and N^{pred}, disaggregated and non-disaggregated forecast, respectively. However, a third case is developed. This is refered as partial disaggregation, $\hat{N}^{pred}_{partial}$, since it uses the basic not-disaggregated forecast, N^{pred}, during the night (18h to 06h) when solar irradiance is zero or very small and the new disaggregated forecast, $\hat{N}^{pred}$, during the day-light hours (06h to 18h) when irradiance is not negligible.

At the end, those comparisons prove the effectiveness of forecasting Net Load by the use and partial use of disaggregation in contrast with not using it.

A. Synthetic PV

The same 10 SS and 5 PV sizes used in Section III-A were used to generate load forecasts. The separate results from each feeder under each DG installed capacity load forecasting were collected and averaged.

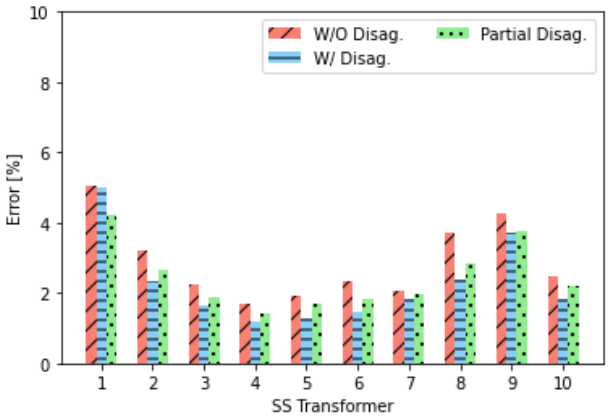


Fig. 8. Effect of Disaggregation on Net Load Forecast Results when Considering Synthetic PV Generation

The previous was done to prove that dissagregation improves the predictions independently of the size of the associated Power Transformer and if disaggregation is used fully or partially. Fig. 8 provides a comparison between the average error in the 10 SS Power Transformers load forecast without, partially and fully using of dissagregation. A constant improvement is seen when using the methodology provided in this thesis. Even though in this case, the largest change was achieve by using all the disaggregated forecast resulting in an improvement of 0.63 pp.

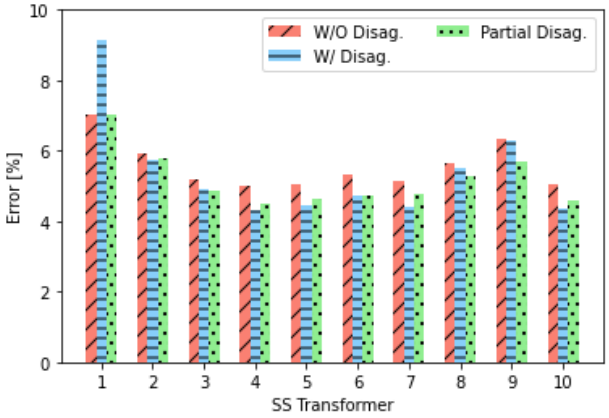


Fig. 9. Effect of Disaggregation on Net Load Forecast Results when Considering Real PV Generation

B. Real PV

By using the same feeders as in Section III-B, load forecasting was run to illustrate the benefit from using disaggregation before predictions are made. The mean errors of the forecast per feeder are presented in Fig. 9.

When considering a more realistic case study, the method that performed better involved the partial use of disaggregation. The change in percentage error was of 0.38 pp.

C. Results Discussion

Looking back at the results from Section II-B, one may see that the PV panel's performance affects the final results. In a perfect information scenario, such as the one we have illustrated using synthetic PV, a 98.82% accuracy was achieved. This contrasts with the 85.86% accuracy seen when using imperfect information from real PV data. Nevertheless, the results achieved will bring value to a DSO in the way that it is avoiding the installation and maintenance costs related to the deployment of sensors across the grid.

Regarding the application of disaggregation to load forecasting, Table IV summarizes the previously presented results. It is clear from such results that, by applying the methodology presented in this thesis, a constant improvement is achieved either by using full or partial disaggregation.

TABLE IV
MEAN ERRORS [%] ON NET LOAD FORECAST

	Without Disaggregation	With Disaggregation	With Partial Disaggregation
Synthetic PV	2.90	2.27	2.45
Real PV	5.57	5.38	5.19

The contrast between errors obtained with synthetic and real PV data deserves some notice. Disaggregation proved to be the most effective method to increase the accuracy of load forecast under the controlled environment in which we used synthetic PV generation. However, when moving the testing procedures to a more realistic environment, disaggregation let to a decrease in net load forecast accuracy in a few SS transformers. Such results led to the implementation of a partial disaggregation approach, in an effort to return to the improvement path found for the previous case study. At the end, either by using disaggregation calculations for daytime only or for the whole day, disaggregation was always capable of decreasing the average error under both scenario used.

After noticing the results on Net Load forecasting presented in Fig. 8 and Fig. 9, a deeper investigation was done. In essence, to look at the separate forecast of the estimated PV Generation, $\hat{PV}$, and the estimated Natural Load, $\hat{L}$. On first place, $\hat{PV}^{pred}$ average errors for the 10 SS Transformers are seen in Fig. 10. Regardless of which scenario is used, the average errors are very low and constant in all the analyzed feeders. However, the same does not happen with $\hat{L}^{pred}$ average errors seen in Fig. 11. There, the errors significantly increase when applying the Real PV scenario and the behavior among the 10 SS Transformers varies as well. That behaviour highly conditions the Net Load predictions observed Fig. 8 and Fig. 9 no matter if disaggregation calculations are not used, partially or fully used. Nevertheless, the previous is only caused by the forecast model used and any improvement in such lies outside the scope of this thesis. What is worth mentioning is that, without taking into account the performance of the forecasting model itself, the use or partial use of the presented disaggregation methodology reduced the average prediction errors of Net Load as seen in Table IV.

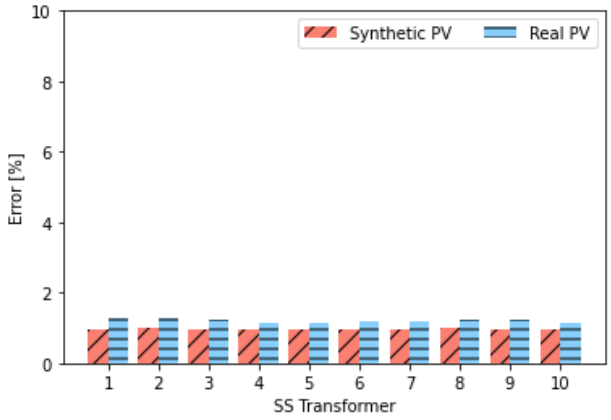


Fig. 10. Estimated PV Generation Prediction, $\hat{PV}^{pred}$, Average Errors under Synthetic PV and Real PV scenario

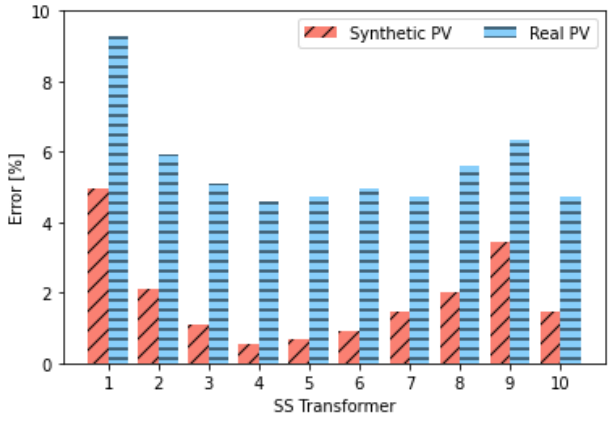


Fig. 11. Estimated Natural Load Predictions, $\hat{L}^{pred}$, Average Errors under Synthetic PV and Real PV scenario

V. CONCLUSION

This thesis developed a non-intrusive methodology to identify distributed generation' installed capacity in a distribution grid by analyzing the corresponding transformer's active power measurements and local meteorological data. The proposed methodology was based on classical mathematical optimization and statistical analysis concepts, avoiding complex learning approaches found in the literature. Results have proven to be very promising: in average, the estimated installed DG capacity will have an accuracy of about 99%, when

considering a Synthetic PV scenario, and an accuracy of more than 85% when considering a Real PV scenario.

The previous shows that the proposed methodology can be considered accurate. However, it was proven that, under realistic scenarios, the bigger the DG installed capacity the more average percent error is expected. Hence, the referred big magnitudes in DG were reaching the associated transformer rated capacities. When a situation like that happens, the DSO will take preventive actions to not compromise the security of the network. Nevertheless, this will be a more common situation in the future.

DSOs and other stakeholders will consider the proposed methodology valuable in the sense that its implementation will avoid high installation and maintenance costs related to the deployment of sensors across the grid. Given that these actors are only aware of Net Load measurements, being able to observe hidden profiles such as the Natural Load and PV Generation profile will allow them to improve their daily basis operations and expansion plans.

After all the performed analyzes on Net Load, Natural and PV Generation profiles, is cleared the big dependence that Net Load has with the always variable PV Generation profile. As the size of the last tends to increase during time, so does its variability according to sun light, and consequently, the frequency of observing behaviours like forward and reverse power flows in the LV distribution grid.

The methodology was develop is such way that it was possible to test its performance under various PV penetrations magnitudes in the LV distribution grid. The previous was accomplished given that the estimated installed DG capacity, X, is a user-defined variable. Thus, the proposed methodology is ready to work with multiple DG capacities.

Furthermore, a practical use for the methodology was proposed as an illustration. Load forecast results were presented to illustrate the advantage of having good estimates of the PV capacity to accurately disaggregate Net Load into Natural Load and PV generation profiles as a prior step before forecasting them. Although results under Synthetic PV scenario were better than those under Real PV scenario, a trend in decreasing the predictions' error value when using full or partial disaggregation in comparison with not using it was clearly achieved in both cases.

ACKNOWLEDGMENT

I want to deeply thank my supervisors, Prof. Pedro Carvalho and Prof. Marcelino Ferreira, for the trust they put on me when accepting me on board of the project at INESC-ID. In addition to Prof. Hugo Morais, their help, comments and guidance throughout the project was invaluable and I am deeply thankful for that. This work was partially supported by Portuguese national funds through Fundação para a Ciência e a Tecnologia with reference UIDB/50021/2020 and by E-REDES, the Portuguese Distribution System Operator.

REFERENCES

[1] T. Aziz and N. Ketjoy, "Pv penetration limits in low voltage networks and voltage variations," *IEEE Access*, vol. 5, pp. 16 784–16 792, 2017.

[2] P. Denholm, M. O'Connell, G. Brinkman, and J. Jorgenson, "Overgeneration from solar energy in california. a field guide to the duck chart," 11 2015. [Online]. Available: https://www.osti.gov/biblio/1226167

[3] R. Bründlinger, "Grid codes in europe - overview on the current requirements in european codes and national interconnection standards," 11 2019.

[4] IREC, "Model interconnection procedures 2019," 09 2019.

[5] K. Cetin, M. Siemann, and S. C, "Disaggregation and future prediction of monthly residential building energy use data using localized weather data network," in *ACEEE Summer Study on Energy Efficient Buildings*, Pacific Grove, CA, 2016, pp. 21–26,.

[6] G. S. Ledva and J. L. Mathieu, "Separating feeder demand into components using substation, feeder, and smart meter measurements," *IEEE Transactions on Smart Grid*, vol. 11, no. 4, pp. 3280–3290, 2020.

[7] E. Vrettos, E. C. Kara, E. M. Stewart, and C. Roberts, "Estimating pv power from aggregate power measurements within the distribution grid," *Journal of Renewable and Sustainable Energy*, vol. 11, no. 2, p. 023707, 2019. [Online]. Available: https://doi.org/10.1063/1.5094161

[8] G. Hart, "Nonintrusive appliance load monitoring," *Proceedings of the IEEE*, vol. 80, no. 12, pp. 1870–1891, 1992.

[9] M. Sun, F. M. Nakoty, Q. Liu, X. Liu, Y. Yang, and T. Shen, "Non-intrusive load monitoring system framework and load disaggregation algorithms: A survey," in *2019 International Conference on Advanced Mechatronic Systems (ICAMechS)*, 2019, pp. 284–288.

[10] W. Kong, Z. Y. Dong, J. Ma, D. J. Hill, J. Zhao, and F. Luo, "An extensible approach for non-intrusive load disaggregation with smart meter data," *IEEE Transactions on Smart Grid*, vol. 9, no. 4, pp. 3362–3372, 2018.

[11] M. Zhuang, M. Shahidehpour, and Z. Li, "An overview of non-intrusive load monitoring: Approaches, business applications, and challenges," in *2018 International Conference on Power System Technology (POWERCON)*, 2018, pp. 4291–4299.

[12] S. S. Hosseini, K. Agbossou, S. Kelouwani, and A. Cardenas, "Non-intrusive load monitoring through home energy management systems: A comprehensive review," *Renewable and Sustainable Energy Reviews*, vol. 79, pp. 1266–1274, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1364032117307359

[13] L. Pereira and N. Nunes, "Performance evaluation in non-intrusive load monitoring: Datasets, metrics, and tools—a review," *Wiley Interdisciplinary Reviews: data mining and knowledge discovery*, vol. 8, no. 6, p. e1265, 2018.

[14] H. K. Iqbal, F. H. Malik, A. Muhammad, M. A. Qureshi, M. N. Abbasi, and A. R. Chishti, "A critical review of state-of-the-art non-intrusive load monitoring datasets," *Electric Power Systems Research*, p. 106921, 2020.

[15] R. L. Burden, J. D. Faires, and A. M. Burden, *Numerical analysis*, 3rd ed. PWS Publishers, 1985.

[16] J. Macqueen, "Some methods for classification and analysis of multivariate observations," in *In 5-th Berkeley Symposium on Mathematical Statistics and Probability*, 1967, pp. 281–297.

[17] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and Édouard Duchesnay, "Scikit-learn: Machine learning in python," *Journal of Machine Learning Research*, vol. 12, no. 85, pp. 2825–2830, 2011. [Online]. Available: http://jmlr.org/papers/v12/pedregosa11a.html

[18] EDP. Sunlab faro - meteo 2017. [Online]. Available: https://opendata.edp.com/explore/dataset/sunlab-faro-meteo-2017/information/

[19] S. Ibrahim, I. Daut, Y. Irwan, M. Irwanto, N. Gomesh, and Z. Farhana, "Linear regression model in estimating solar radiation in perlis," *Energy Procedia*, vol. 18, pp. 1402–1412, 2012.

[20] EDP. Sunlab faro - pv 2017. [Online]. Available: https://opendata.edp.com/explore/dataset/sunlab-faro-pv-2017/information/

[21] S. Kay and S. Marple, "Spectrum analysis—a modern perspective," *Proceedings of the IEEE*, vol. 69, no. 11, pp. 1380–1419, 1981.

[22] B. Zdaniuk, *Ordinary Least-Squares (OLS) Model*. Dordrecht: Springer Netherlands, 2014, pp. 4515–4517. [Online]. Available: https://doi.org/10.1007/978-94-007-0753-5$_2$008