Electrical Power and Energy Systems 136 (2022) 107653

Contents lists available at ScienceDirect

International Journal of Electrical Power and Energy Systems

journal homepage: www.elsevier.com/locate/ijepes

Forecasting voltage harmonic distortion in residential distribution networks using smart meter data

Pablo Rodríguez-Pajarón [a,*], Araceli Hernández Bayo [a], Jovica V. Milanović [b]

[a] Escuela Técnica Superior de Ingenieros Industriales, Universidad Politécnica de Madrid, Madrid, Spain
[b] Department of Electrical and Electronics Engineering, The University of Manchester, Manchester, UK

ARTICLE INFO

Keywords:
Distribution network
Neural network
Power quality
Smart meter
Voltage distortion

ABSTRACT

This paper introduces a methodology to forecast voltage total harmonic distortion (THD) at low voltage busbars of residential distribution feeders based on the data provided by a limited number of smart meters. The methodology provides relevant power quality indices to system operators using only the existing monitoring infrastructure required for demand response operation. Different algorithms for voltage THD forecasting are implemented, including artificial neural networks, and their performance is tested and compared. The necessary coverage of smart meters for the acceptable accuracy of the estimated THD is also established. The estimation algorithms are validated considering probabilistic demand load model developed based on typical harmonic injections of household devices obtained from measurements and using a typical European low voltage test-feeder with 471 residential consumers.

1. Introduction

In recent years, technological advances, very often driven by energy saving policies, have encouraged a rapid increase in the use of single-phase household appliances based on power electronics. Motors integrated in domestic devices are being increasingly equipped with variable speed drives and traditional lighting equipment is being progressively replaced by Compact Fluorescent Lamp (CFL) and Light Emitting Diode (LED) technology. As an example, only in the last decade, CFL have replaced around 25% of conventional lamps [1,2].

The collective effect of these dispersed non-linear (NL) domestic loads has led to observe harmonic current distortion levels in residential feeders higher than those of the industrial and commercial feeders [3]. Hence, excessive voltage harmonic distortion in residential distribution systems is a growing matter of concern that utilities will increasingly have to cope with in a near future. Maintaining harmonic levels under established limits is a joint responsibility involving both, end-users and system owners or operators and, consequently, harmonic limits are recommended for load currents and also for system voltages [4–6].

In order to anticipate and mitigate problems that arise due to harmonic presence, utilities must be able to predict the expected harmonic impact and to evaluate the ability of existing distribution networks to accommodate new NL loads. One possible way to assess harmonic levels is through extensive monitoring programmes, however, this is a costly solution in terms of effort and financial resources required if applied at low voltage (LV) buses. Another approach is to perform harmonic simulations by modelling the distribution network and representing the loads as stochastic individual harmonic sources [7–10] or aggregated harmonic loads [11–13]. The accuracy of simulation based approaches, however, depends on the accuracy of the mathematical models used and the validity of the considered scenarios.

More recently, the approaches that extend conventional state estimation methodology to harmonic estimation at distribution systems have been proposed. They are based on the integration of Phasor Measurement Units (PMUs) at the distribution level in order to locate harmonic sources [14–16].

In this paper, a new way of utilising intelligent electronic devices installed in the smart distribution networks is explored in order to gather information about power quality (PQ). In particular, the proposed methodology enables the prediction of voltage Total Harmonic Distortion (THD_V) at LV busbars of a residential feeder based on a reduced number of measurements provided by smart meters with sub-metering capabilities, i.e., smart meters that are capable of recording power demanded by the individual home appliances. These meters are usually

* Corresponding author at: Escuela Técnica Superior de Ingenieros Industriales, Universidad Politécnica de Madrid, Calle José Gutiérrez Abascal 2, 28006, Madrid, Spain.
E-mail address: pablo.rpajaron@upm.es (P. Rodríguez-Pajarón).

https://doi.org/10.1016/j.ijepes.2021.107653
Received 2 October 2020; Received in revised form 17 August 2021; Accepted 20 September 2021
Available online 7 October 2021

intended to perform decomposition of demand into load categories meaningful for advanced demand side management. Such demand side techniques are recognized as one of the key options to achieve a better operation of the electric systems in future distribution networks [17]. As an example, in the U.S., it was estimated that the participation of residential customers in demand response (DR) might bring up to half of the total peak reduction [18,19]. The network operator in the U.K. is planning to obtain 30 to 50% of its balancing services through DR, which will only be possible if enough load monitoring devices are installed [18,20,21]. By means of the proposed approach, both, development of advanced DR strategy and PQ estimation and control can be achieved in a cost effective way by using a single measurement device, i.e., smart meters with sub-metering capabilities. This approach is inspired by the existence of numerous trial sites in residential sector with such smart meters [22–24].Since DR has been recognized as one of the most cost-effective options to operate the system, the deployment of smart meters with submetering capabilities is expected to increase to enable a better observability of the ends users' behavior and their potential to participate in the network daily operation [18,25]. In addition, there are some metering designs which propose using the so-called global Home Management Systems, which not only measure power consumed by each load in the dwelling, but also allow controlling its demand [26,27]. An alternative approach to submetering is performing demand decomposition based on non-intrusive load monitoring methods. These methods for load decomposition have been extensively analyzed in the literature [23,28–31].

To obtain an adequate performance of a DR programme relaying on smart meters with submetering functionality a penetration level between 5% to 20% of submetering enabled meters is needed [18] which is similar to the smart metering penetration level proposed here. One of the key objectives of this study is to provide the accurate information about PQ indices to DSO without installing any additional monitors and to rely only on the information provided by the existing smart meters with sub-metering functionality. The results show that a reasonably small number of end-users are required to have this type of smart meter in order to estimate sufficiently accurately the THD_V at LV busbars avoiding the need to install specific PQ meters and taking advantage of the monitoring infrastructure already installed.

Therefore, the proposed methodology enables the forecast of present or future harmonic scenarios at LV networks by integrating new functionalities into already existing monitoring devices. Artificial neural networks (ANNs) are used to fulfil this aim. The cases tested in this paper shows that harmonic distortion can be predicted at different buses, from the secondary of the main transformer to the farthest bus along the feeder, which provides a useful tool to monitor power quality behaviour of the network or even deciding the adequacy of connecting new loads. The main contributions of this study include: i) The artificial neural networks based methodology for forecasting time varying THD_V at LV busbars by using measurement data provided by smart meters with sub-metering capability installed at residential properties without the need of installing dedicated PQ meters. ii) The assessment of the percentage of smart meters with sub-metering capability required to estimate THD_V with a certain confidence level. iii) The comparison of different algorithms performance for forecasting THD_V based on the data provided by smart meters.

2. Methodology overview

2.1. THD forecasting methodology

Voltage harmonic distortion at residential networks is directly related to the currents injected by NL household devices [3]. These devices typically behave as a highly variable and unpredictable electrical demand due to their stochastic connection and disconnection to the network [32]. Consequently, harmonic currents injected by residential loads present also a stochastic behavior and are hardly foreseeable for an individual consumer [33]. Nevertheless, the aggregated demand of several consumers has a more predictable pattern. The decomposition of this aggregated demand into different load categories has been explored mainly for demand side management purposes [18,20]. In this approach, household appliances are classified first as linear or NL loads. Next, four different categories of NL loads are established according to their characteristic harmonic current spectrum: (i) NL lighting (mainly CFLs and LEDs), (ii) Switch Mode Power Supply (SMPS) (i.e., domestic electronic devices), (iii) variable speed drives (e.g., washing machines or modern fridges and heat pumps) and (iv) directly connected induction motors, i.e., IMs, (e.g., water pumps or old fridges and washing machines) [34,35,11].

Taking into account this decomposition of aggregated residential demand, the general methodology proposed in this work is based on two steps:

(i) Step 1: The decomposition of the total power demand into the power demand by individual NL load categories using ANNs.The data from a small amount of smart meters with submetering capabilities are used to feed ANN, which are then trained with MATLAB ANN tools [36].

(ii) Step 2: Estimation of the THD_V at LV busbars based on the relationship between THD_V and the predicted individual demand of different NL load categories (several methods are compared and discussed in Section 4 in order to select the most appropriate one for this estimation).

This general approach is depicted in Fig. 1. Step 1 (demand decomposition into NL categories) is described in detail in Section 3 while step 2 of the methodology (THD_V prediction) is explained in Section 4.

ANNs are applied in both steps of Fig. 1. Different alternative machine learning tools and methodologies could be applied in these steps, including those known as Deep Neural Networks (DNN). However, this work primarily focuses on demonstrating the feasibility of application of artificial intelligence techniques for estimating PQ parameters with the information needed for DR. Performing an exhaustive comparative analysis of multiple artificial intelligence algorithms and structures to solve the stated problem is beyond the objectives of this paper.

2.2. Underlying assumptions

The formulation of the proposed methodology considers the following assumptions:

(i) Step 1: Decomposition of total power demand into NL categories. The following assumptions have been made for the application of the step 1 of the method:

- A certain number of smart meters with sub-metering capabilities are assumed to be installed in the network. The number of clients with a meter is discussed in Section 3. The accuracy of performance has been tested and the accurate performance has been reached with 10% of clients with such meters.
- The location of meters is arbitrary, and it is not selected according to any specific strategy.

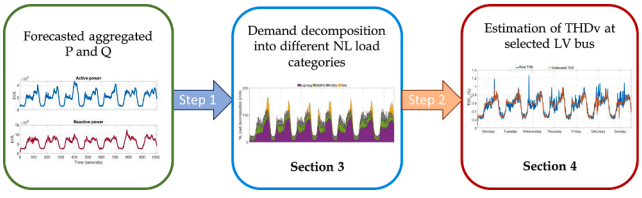

Fig. 1. Overview of the proposed THD_V forecast methodology.

- Power measured by these smart meters is known during 10 to 20 weeks in order to train ANN networks.
- The total aggregated demanded power at the secondary of transformer is also known.
- Four NL categories of loads are used for demand decomposition: NL lighting, Switch Mode Power Supply (SMPS), variable speed drives and directly connected induction motors [34].
- Network parameters are not known.
- Ten-minute intervals are used for power measurement and decomposition in accordance with the requirements of EN50160 [4].

The output of this step is the forecast of the power demanded by each NL category for each 10-min interval.

(ii) Step 2: THD_V forecast. In the second step, the following assumptions are considered:

- Load decomposition estimated in step 1 of the method is used.
- A PQ monitor capable of recording THD_V had been previously trained during 10–20 weeks.
- Network parameters are not known.
- Ten-minute intervals are used for THD_V training and prediction.

The output of this step is the forecast of THD_V in each 10-min interval.

3. Demand decomposition into NL load categories

Only devices that inject distorted current are of interest for the purpose of forecasting THD_V at LV busbars. Therefore, prior to perform classification into NL load categories, residential loads have to be classified first into linear and NL, i.e., the ones characterized by distorted current.

3.1. Linear and NL loads

In order to classify linear and NL loads, the typical value of total current harmonic distortion (THD_I) [5] of the most common household appliances has been probabilistically characterized. THD_I is calculated by means of (1):

$$THD_I = \sqrt{\sum_{i=2}^{H} \left(\frac{I_i}{I_1}\right)^2} \times 100 \quad (1)$$

where i stands for each harmonic order considered up to the highest order H, I_i is the injected harmonic current and I_1 is the fundamental current.

Loads characterized by current harmonic injections with very low THD_I can be considered as non-distorting loads (i.e., linear loads) while higher levels of THD_I are typical of NL loads . However, the threshold up to which THD_I is high and thus the load is considered NL is not clear.

In order to set a threshold value of THD_I that can discriminate linear and NL loads, the following process has been followed:

(i) Demand curves of several houses have been simulated with the probabilistic tool from Crest model [37] during 20 weeks. Crest tool is a model to synthetically generate demand profiles of a single household, based on probabilistic values (such as active occupancy of the house, temperature, sun irradiance and time of use of appliances). This model follows a bottom-up procedure, i.e., residence demand is computed by adding each device's demand, minute by minute.

(ii) A current injection spectrum has been allocated to each household device. Current injections are generated following a bottom-up approach, based on the demand model developed by Crest [37] and real, measured harmonic spectrum reported in Panda database [38]. Harmonic injections depend on the ON-OFF operation state of residential loads, which is highly stochastic. The load operation is hence, simulated based on probabilistic demand models reported in [37]. Once the operation state of the load is estimated, harmonic currents injected in the ON state are synthetically generated (with both, magnitude and phase angle) according to characteristic multivariate gaussian distributions of the harmonic current injections of each device. These characteristic gaussian distributions have been identified for each residential appliance by performing cluster analysis techniques [39] on the set of corresponding, measured harmonic currents for different devices available in Panda database [38]. In this way, the variability in the harmonic injection by different devices of a certain appliance type is included in the model. For devices with very little data available in Panda, additional measurements have been performed.

The simulation of the random operating state of loads, together with the probabilistic characterization of harmonics currents injected by devices, allows a stochastic modeling of harmonic injection of residential loads. This procedure follows a similar approach reported in [7,9].

With this, the harmonic currents injected by the devices and, therefore, by the residence can be determined.

(iii) These generated current harmonic profiles have been applied to a small network [40], where the harmonic injection problem has been solved for every minute of the week. This calculation has been done using the OpenDSS [41] simulation tool, by modelling residential loads as Norton equivalents [7]. Through harmonic injection analysis, harmonic voltages are obtained, allowing to compute THD_V.

(iv) A regression has been calculated between the THD_V obtained at the secondary side of the main transformer at each minute of a week and the percentage of power demanded by NL loads in that minute, when considering different THD_I thresholds to discriminate linear from NL loads. Fig. 2 shows that these two magnitudes (THD_V at secondary of transformer and proportion of NL power with respect to the total demanded power) have higher correlation index and regression slope when THD_I threshold is over 15%. This study has been done for 20 different weeks, each represented by a dot in Fig. 2.

The results indicate that loads with less distorted current show lower influence i.e., lower correlation index on voltage distortion at LV busbars. Therefore, in this work, the loads whose harmonic spectrum leads to a THD_I>15% are designated as NL loads. Similar results have been found when applying exactly the same process described here to a different, larger, test feeder.

Taking into account the variability among a particular type of devices and the probabilistic approach followed to model their harmonic spectrum, some devices can be classified as NL at some houses and linear at others. Table 1 shows some of the most typical household appliances

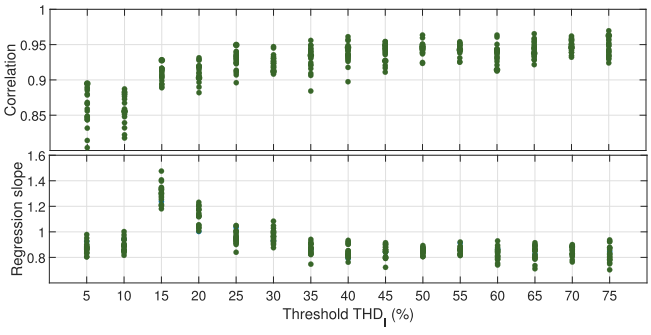

Fig. 2. Correlation between THD_V at secondary of transformer and NL load proportion at 55 load network [40], with different THD_I thresholds.

Table 1
NL devices for a sample of appliances in 500 aggregated dwellings for a summer week.

	Participation in energy consumption (%)	% NL	Category
Resistive devices (oven, kettle, e-heating)	18.8	0	Linear
Washing machine	18.7	75.4	VSDs and IMs
TV	15.8	60.8	SMPS
Fridge and freezer	14.6	40.4	VSDs and IMs
PC	7.5	100	SMPS
Lighting	6.0	58.8	NL lighting
Vacuum	2.1	33.1	VSDs
Rest (microwave,dish-washer, etc.)	16.5	10.0	SMPS

obtained for 500 houses. For this study, it has been assumed that out of all lighting devices 41% are linear (halogen or incandescent lights) and 59% NL, with LEDs and CFLs representing 24% and 35% of the total, respectively [1]. The second column of Table 1 shows the percentage of the energy consumed by a certain type of appliance during a summer week obtained using Crest tool [37]. The third column indicates the percentage of devices of a given type classified as NL (i.e., $THD_I > 15\%$) based on probabilistic characterization of their harmonic spectrum. For example, based on Table 1, TVs consumed 15.8% of the energy and 60.8% of the TV devices were classified as NL. Therefore, during the week, assuming all the TV devices have the same or very close characteristics, $15.6\% \times 60.8\% = 9.6\%$, i.e., approximately 9.6% of the energy consumed by residences has non-linear characteristics corresponding to the spectrum associated with TVs.

To obtain values of Table 1 Panda current harmonic measurements [38] were allocated to Crest demand model [37] as previously described in step ii).

The assumed harmonic current injections have been considered independent from voltage supply distortion [7]. Possible variations in current injections caused by voltage supply distortion can be considered as part of the uncertainties included in the assumed probabilistic model.

3.2. Characteristics of decomposed NL demand

Fig. 3 shows the aggregated power demand of 500 individual residences in a summer and winter week. This demand has been synthetically obtained using Crest tool [37]. With the aim of observing its characteristic behaviour, the aggregated demand has been decomposed into linear and NL demand, according to the THD_I of the appliances that contributed to it and according to the defined THD_I threshold and by means of the method previously described. In addition, NL demand has been decomposed into four categories depending on the type of household appliance that causes such demand. These categories (described in Section 2) are: (i) lighting NL lamps, (ii) SMPS, (iii) VSDs and (iv) IMs.

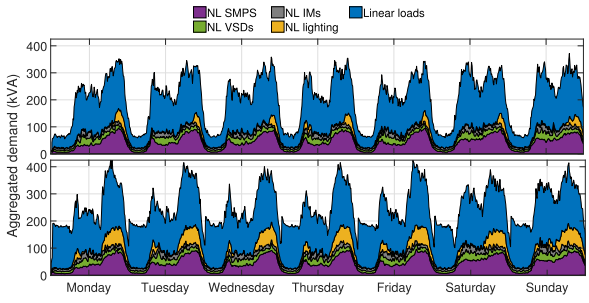

Fig. 3. Aggregated NL demand decomposition of 500 residential consumers in a summer (top) and winter (bottom) week.

As it can be seen from Fig. 3, both, the total aggregated power demand, and the aggregated power demand of NL loads follow a typical daily trend. It is this repetitive behavior of decomposed demand shown in Fig. 3 which can facilitate its prediction.

In this approach, ANNs are used to forecast decomposition of residential loads into the specified NL load categories, based on information of the total aggregated active and reactive power demand.

3.3. NL demand forecast

The general process to forecast demand decomposition is depicted in top of Fig. 4. Neural networks are trained with the past information of load disaggregation provided by smart meters installed in a certain percentage of houses as a realistic scenario in the future smart distribution grid [18,19]. These submetering-enabled smart meters are assumed to be installed for DR management in distribution networks, therefore their location has been randomly selected and has not been discussed in this paper as it is very unlikely that the placement of these meters in real networks will be guided by any optimization procedure related to harmonic monitoring, or even the demand response programmes, in the future.

In order to train the neural networks, it is assumed that smart meters record real power of individual appliances in the houses where they are installed. According to this assumption, smart meter recorded demand from individual appliances can be aggregated at the concentration point following the proposed load decomposition into NL loads categories. The ANN is trained during several weeks with the total demand and the demand recorded for each load category (top-left of Fig. 4). Once ANN is trained, forecasts of demand decomposition of future weeks can be made just by providing the forecast of total power consumption (real and reactive) by all residential consumers (top-right of Fig. 4) at the same substation. In this study, demand is calculated with a 10 min interval resolution, in accordance with the time step required by power quality standard EN50160 [4] and based on [42] where it is reported that demand can be measured over periods from 1 to 60 min.

Simulations carried out in these studies show a very small influence of the background harmonic voltage distortion at the MV level on the predicted levels of distortion at LV buses, so this effect has been neglected in the cases studied. In any case, considering that MV voltage background distortion tends to follow a cyclic behaviour [43], its summation effect could be easily incorporated in ANNs training if required.

3.4. ANN architecture for load decomposition

In this study, two different neural networks are tested for the process

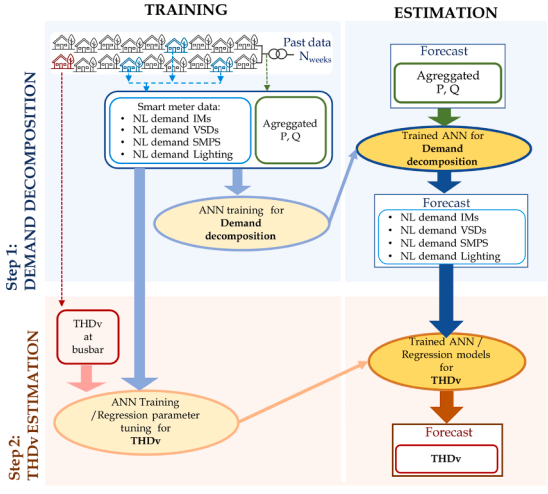

Fig. 4. THD_V estimation methodology.

described in top of Fig. 4 considering different percentage of smart meters coverage in the grid. First neural network is a two-layer feed-forward neural network and trained with Bayesian Regularization Backpropagation algorithm [44,45]. The input layer has two neurons (active and reactive power) and the output has four (one for each non-linear load category). Transfer functions of hidden and output layer are log-sigmoid and tan-sigmoid, respectively, following the approach used in similar applications [21,46]. The hidden layer has 32 neurons, as suggested in [21,46] and finally decided on through a trial-and-error procedure as a compromise between the accuracy of assessment and the computational time [39].

The other neural network used takes into account the time-series characteristic of data to be forecasted (in this case, the demand per NL category). It is a nonlinear autoregressive with external input neural network [47], with similar architectures for the input, hidden and output layers. Both networks have been adopted from [36], applied in MATLAB. Neural network architecture and parameters tested in this study are adopted from [18–21] and found to be a good compromise between the training requirements and accuracy. Other ANN architectures or parameters could be used, and the same methodology followed.

Training and accuracy of ANN disaggregation algorithms are exhaustively studied in [18,21]. Non-drastic changes in loads can be handled with sufficient robustness and accuracy by the trained network.

4. Estimation of THD_V at LV busbars

The decomposition of demand described in previous section established four categories of NL loads that are characterized by their harmonic current spectrum and that determines voltage harmonic distortion at LV busbars. Therefore, the second step of the proposed method is the prediction of voltage THD_V at LV busbars by using the estimation of demand of the defined NL load categories. This process is illustrated at the bottom of Fig. 4.

In this work, four different algorithms are proposed and tested in order to perform the THD_V prediction:

(i) Feed-forward neural network [21,48]. The input layer of this ANN has 6 neurons (P, Q and the demand estimated for the four NL categories previously defined) and the output layer has one (THD_V). Therefore, once trained during a certain number of weeks the ANN can provide a forecast of THD_V.

Feed-forward neural network applied here is trained with Bayesian Regularization Backpropagation algorithm [44]. Transfer functions of hidden and output layer are, again, log-sigmoid, and tan-sigmoid respectively. Number of neurons in the hidden layer was calculated according to [21,46] and then corrected in a trial-and-error procedure to 57, which has been found the most efficient.

(ii) Autoregressive neural network [47,48]. Similarly to the ANN used before, its input layer has 6 neurons (P, Q and the demand estimated in the four NL categories) hidden layer has 27 and output one (THD_V) neuron. Transfer functions for hidden and output layers are, too, log-sigmoid and tan-sigmoid. As in previous ANN introduced, other architecture and parameters could be applied following the same methodology.

(iii) Simple regression [39,49]. A simple regression between the total aggregated NL power (that is the sum of the four proposed categories of NL demand) and THD_V is calculated. Therefore, in this method, the input is only the total percentage of NL demand and disaggregated values of NL categories are not used. Tuning of the regression parameters (correlation slope and intercept) is made during training weeks, when both NL demand and THD_V are supposed to be available. Regression is computed with least squares fitting algorithm.

(iv) Multivariate regression [39,49]: This is an evolution of the former, NL demand for each of the four load categories are used as predictors for a multivariate regression. Regression is computed again with least squares fit algorithm. NL demand of each category are the inputs to this algorithm. Regression model is tuned with past data of NL disaggregated demand and THD_V at the considered bus.

The complete methodology flowchart is included in Fig. 4, where all the inputs and outputs of the methodology steps are included. Note that regression algorithms are only proposed for the independent estimation of THD_V while the ANNs (with different architectures) will be used for both, the estimation of THD_V and for demand composition.

5. Performance of the proposed methodology

Two different ANNs have been tested to accomplish the first step of the methodology (NL demand decomposition estimation) and four algorithms have been proposed for the estimation of THD_V. Based on independent evaluations of their accuracy the most favorable combination of the different algorithms have been tested and used for the illustration of the overall process of the estimation of THD_V.

Fig. 5 shows the procedure followed to illustrate and analyse the performance of the algorithms proposed in this paper. The validation of step 1 of the method (decomposition of demand) is explained in Section 5.1. The validation of step 2 (estimation of THD_V) is described in Section 5.2. Finally, the overall procedure is analyzed and validated in Section 5.3.

5.1. Demand decomposition using different ANNs

As explained in Section 3.4, prediction of demand decomposition can be performed by means of feed-forward ANN or by autoregressive ANN. Both methods have been implemented for the decomposition of the aggregated consumption of 471 residences. The demand model given in [37] has been applied to generate residence load profiles over 20 weeks for training and testing the ANNs. In order to generate the residence demand profiles, occupancy statistics from Spain have been adopted (25% residences with one resident, 30% with two and 20% with three residents, and 25% with four or more) [50]. These profiles provide the decomposed demand that can be assumed as "real" to be compared with the estimation. The total active and reactive demanded power is also obtained and it will be used as an input to the estimation algorithm (see bottom left of Fig. 5).

In order to assess the required coverage of consumers with smart meters, the performance of the estimation algorithms has been tested for eight levels of smart metering penetration: 1% coverage (that is 5 dwellings out of 471 have submetering capabilities), 5% coverage (24 dwellings), 10% (47 dwellings), 20% (94 dwellings), 40% (188 dwellings), 60% (282 dwellings), 80% (376 dwellings) and 100% (all the 471 dwellings).

In order to quantify the performance of both algorithms the actual real power demand by each NL load category is compared with the power estimated by the algorithm for each load category at the aggregation point. Therefore, the absolute factor error (*AFE*) is calculated at each time interval (10 min) using the Eq. (2), [18]:

$$AFE = |P_{ANN} - P_{real}| \tag{2}$$

where P_{ANN} is the estimated active power demanded by a certain NL load category at a considered time interval and it is determined by the ANN estimation, and P_{real} is the actual power consumed by the NL load category at the same time interval. Both values are expressed in p.u. based on the average aggregate active power at the secondary of transformer during the week, i.e., 190 kW in this study.

The accuracy of estimation of the share of all NL loads (independently of their category) is shown in Fig. 6 for different smart metering coverage and for both ANN. It can be seen that above 10% of smart-

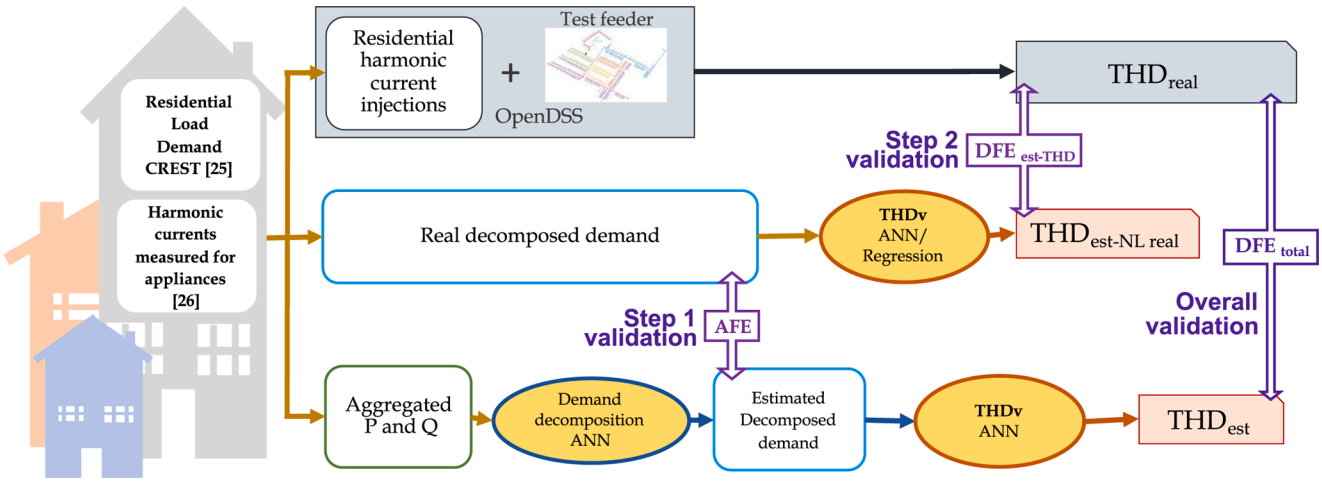

Fig. 5. Validation strategy for steps 1 and 2 and overall method.

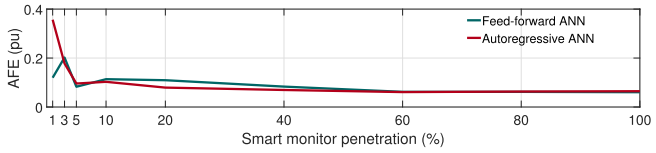

Fig. 6. AFE with 90th percentile confidence level of NL power demand (all load categories) obtained with feed-forward and autoregressive ANNs with different smart meter penetration in the grid.

metering coverage the error is smaller than 0.12 p.u. with both ANNs (i.e. 12% of the weekly average load) in 90% of the time intervals of the week (90th percentile of AFE). It can be also seen that increasing the smart metering coverage above 60% has a very reduced effect on accuracy.

In the left part of Fig. 7 the 90th percentile AFE of the four different load categories is presented in the form of a bar chart for a range of different smart meters coverage when feed-forward ANN is applied for the estimation of demand decomposition. The right part of Fig. 7 shows a similar result when the autoregression is used. In general, both ANN have similar performance with slightly better accuracy of autoregressive ANN for higher smart-metering coverage. However, if monitoring penetration is very small (under 3%) autoregressive ANN shows bigger error.

Fig. 8 shows the 90th percentile AFE obtained with both proposed ANNs for different number of training weeks. In all the cases, 10% monitor coverage is assumed. The performance of the demand decomposition algorithm does not improve substantially above 12 training

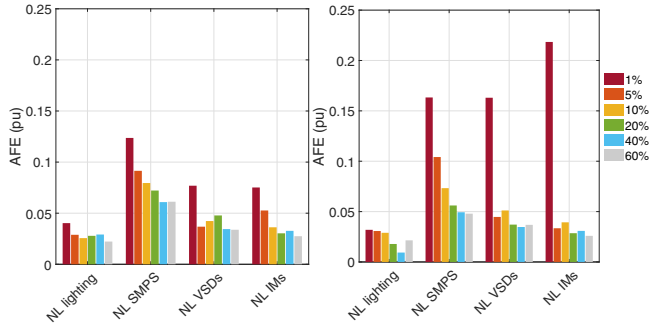

Fig. 7. AFE with 90th percentile confidence level for different categories of NL loads, obtained with feed-forward ANN (left) and autoregressive ANN (right) and different smart meter penetration in the grid.

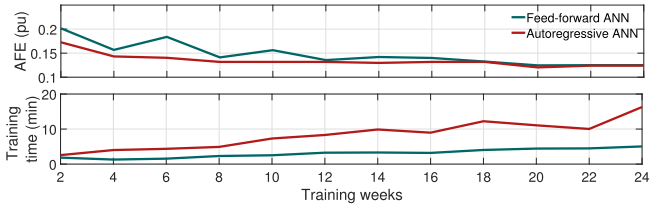

Fig. 8. AFE with 90th percentile confidence level of NL power demand (all load categories) for different number of training weeks ANN training time.

weeks irrespectively which ANN is used. The time required for training the networks with an Intel Core i7-7700 CPU @ 3.6 GHz-32 GB RAM computer is also shown in the graph. It can be seen that it increases linearly with the number of weeks with higher slope in the autoregressive ANN.

The performance of the ANNs shown in Figs. 6 and 8 corresponds to the 90th percentile of the AFE in the prediction of all the 10-min time intervals of a single week. In order to illustrate the performance for different weeks, Fig. 9 shows the complete Cumulative Distribution Function (CDF) of 24 weeks with feed-forward and autoregressive ANN. In all the cases the prediction algorithm has considered 20 training weeks and 10% smart metering coverage. It can be seen that there is a certain variability in the errors among different weeks which is narrower in case of the autoregressive network. It can be seen that during 90% of the time the AFE is below 0.157 for feed-forward ANN and below 0.141 for autoregressive ANN.

Demand decomposition to NL load categories is shown for an example day in Fig. 10. The actual decomposed demand of 471 residential consumers is compared with the decomposed demand estimated by the feed-forward and autoregressive ANNs (10% smart metering coverage and 20 training weeks have been considered).

5.2. THD_V estimation using ANNs and regression algorithms

The second step of the algorithm, shown in Figs. 1 and 4 is based on the estimation of THD_V at LV busbars based on previously estimated decomposition of demand.

To validate this step of the methodology, a test feeder and conventional harmonic power flow calculations are used. Note that the test feeder and harmonic power flow are only used for validation, as the proposed methodology does not make use of any network parameter, since their influence is already integrated in the ANN training. The selected network for this illustration is Representative European Network #7 [51]. It consists of 471 dwellings distributed over seven

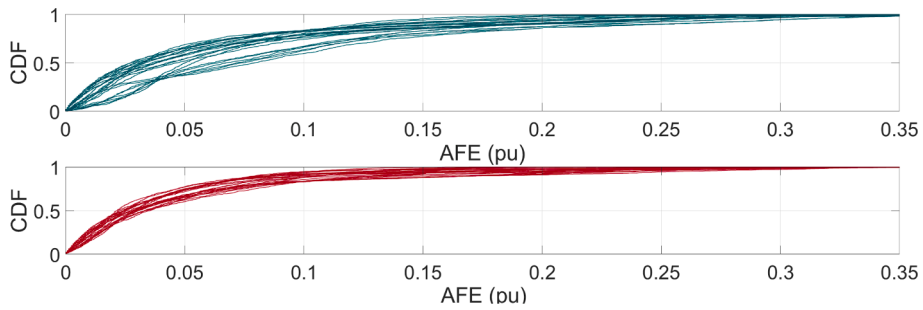

Fig. 9. Cumulative distribution function of AFE of NL power demand (all load categories) with feed-forward ANN (top) and autoregressive ANN (bottom).

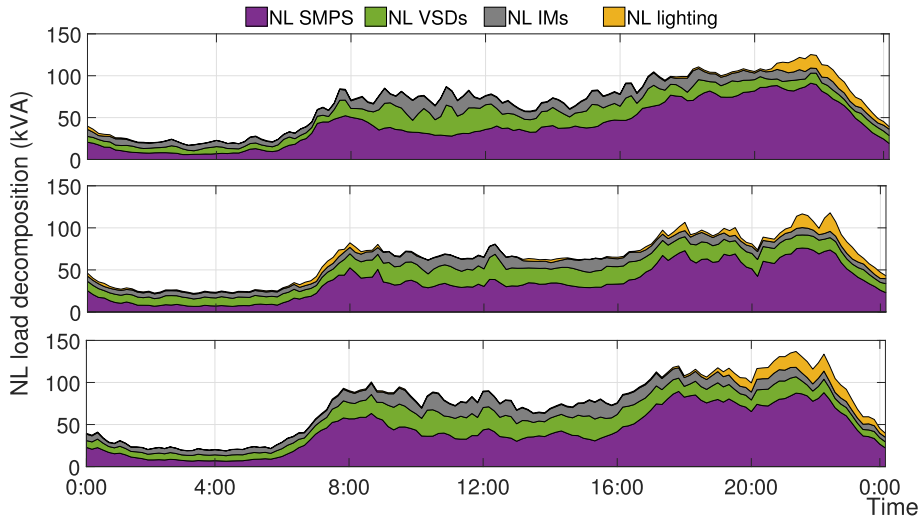

Fig. 10. Demand decomposition: actual values (top) and estimated with feed-forward ANN (middle) and with autoregressive ANN (bottom).

400/230 V radial feeders [51]. Most of them are single-phase and distributed along three phases of the network. All network parameters used in OpenDSS simulations, such as transformer or line parameters or network topology, can be found at [51].They are not included in the paper due to space limitations. In order to better replicate low voltage distribution systems in Europe and to adapt the relation between the rated transformer power and the peak load to realistic values, original [51] transformer rated power has been reduced. With the modified rated power value, the simulated peak load is approximately 70% of the transformer rated power [52]. This adjustment was done as in the original test feeder the transformer rating was 800 kVA while the peak demand on the feeder was around 350 kVA, resulting in unrealistically oversized transformer.

The THD_V can be forecasted at any bus of the network by applying the proposed methodology. In order to show the performance of the methodology, two representative buses are selected to illustrate the results: the secondary of the main transformer of the network and the load bus 471 (marked TR and LB respectively in the single line diagram of Fig. 11).

The validation procedure has been designed so that the performance of all the proposed algorithms for THD_V estimation can be compared without being influenced by the inaccuracy in the previous step of NL

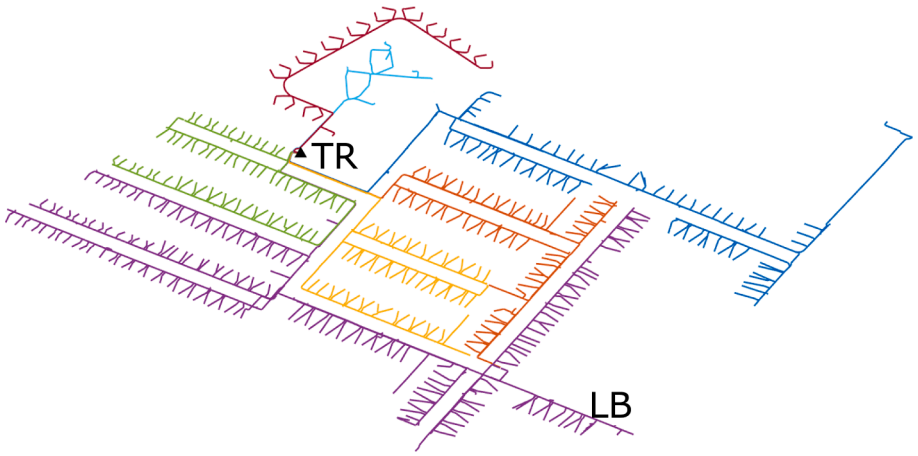

Fig. 11. Single line diagram for Representative European Network #7 [51] with its 7 feeders in different colors.

demand decomposition (see Fig. 5). Taking this into account, the estimation of THD_V is validated according to the following process:

- The demand of the individual residence included in the network of Fig. 11 is generated using CREST tool [37]. Inputs into the model of each of the residence profiles (number of inhabitants, appliances in the dwelling, powers or time of use) are set to be random [37]. All considered scenarios are generated considering the summer season.
- The harmonic spectrum of each household appliance contributing to the demand is generated considering probabilistic distributions of measured real harmonic injections provided in PANDA [38] as explained in Section 3.1.
- Power of the NL loads is aggregated into the defined NL load categories at the substation aggregation point. This aggregation is assumed to be known in the week of prediction instead of being calculated as described in Section 3.3. In this way, the calculated error is caused only by the estimation of THD_V itself, and the different proposed algorithms can be tested without being influenced by the errors in the demand prediction.
- The THD_V estimated at a LV busbars based on the actual demand decomposition ($THD_{est-realNL}$) is compared with the actual THD_V calculated at this LV busbar by the harmonic injection of all individual loads connected to the network (THD_{real}), simulated as Norton equivalents [53] in OPENDSS simulation software. OPENDSS, in its default mode, considers all injecting loads as decoupled Norton equivalents, consisting on a current source and a shunt admittance which demands rated power at fundamental rated voltage. As previously mentioned, network parameters are only required to obtain the value THD_{real} that will be used to validate the method and assess its accuracy.

In order to quantify the error of the estimation, the distortion factor error $DFE_{est-THD}$ is defined as:

$$DFE_{est-THD} = THD_{real} - THD_{est-realNL} \tag{3}$$

where THD_{real} and $THD_{est-realNL}$ were previously defined.

All these steps have been summarized in Fig. 5..

Fig. 12 shows the CDF of the absolute value of $DFE_{est-THD}$ with the four algorithms proposed for THD_V estimation over one week period at bus TR, as representative busbar. In order to make the results comparable, all the algorithms consider the same information, i.e. the same 20 weeks for training the feed-forward and autoregressive ANNs and for obtaining the regression parameters (correlation slope and zero intercept). All the algorithms are also applied to predict the THD_V in the same week. It can be seen that feed-forward ANN and autoregressive ANN perform better than simple and multiple regression algorithms.

Fig. 13 shows the CDF of the error predicted over the period of 24 weeks in order to compare the performance of the prediction algorithm for different weeks at busbar TR. Similar results and variability for both ANNs can be observed as in the previous case. It can be seen that during 90% of the time the error $DFE_{est-THD}$ is below 0.146 for feed-forward ANN and below 0.134 for autoregressive ANN.

5.3. Integrated estimation of the THD_V

In previous sections, the performance of the proposed algorithms was compared independently for the two steps of the process: NL demand decomposition estimation (step 1) and THD_V prediction (step 2). In this section, the whole estimation process is tested using the algorithms which have shown better performance at individual stages, i.e., autoregressive ANN for demand decomposition estimation and both, feed-forward and autoregressive ANN, for THD_V estimation.

Figs. 14 and 15 show a comparison of the real and estimated THD_V obtained using these algorithms, for secondary of transformer (busbar TR) and selected load bus (busbar LB), respectively. In all the cases, 20 training weeks and 10% smart-metering coverage are used as parameters that provide a reasonable accuracy with a good compromise between error and required resources.

As it can be seen in Figs. 14 and 15, estimation shows similar performance for both representative buses. Simulations have shown that similar accuracy is obtained for any other bus in the LV network.

Fig. 16 shows the boxplot of the THD_V values obtained for all 10-min time intervals of a week. It can be seen that both ANNs have a reasonable accuracy in the estimation of the median, the 25th and the 75th percentile. Autoregressive ANN is more accurate in estimating the minimum.

The error of the overall estimation of THD_V, DFE_{total}, is calculated as:

$$DFE_{total}(\%) = THD_{real}(\%) - THD_{est}(\%) \tag{4}$$

where THD_{est} (expressed in %) is the THD_V estimated at the LV busbar following the estimation process and THD_{real} (expressed in %) is the real THD_V obtained from the actual harmonic injections following the process described in Section 5.2.

The error defined in (4) can be split into two components: i) the component of the error made in the estimation of THD_V assuming perfect knowledge of NL decomposition ($DFE_{est-THD}$) given by (3), ii) the component of the error in THD_V prediction caused by the inaccuracy on the estimation of the decomposed NL load categories (DFE_{est-NL}) given by:

$$DFE_{est-NL} = THD_{est-realNL}(\%) - THD_{est}(\%) \tag{5}$$

where $THD_{est-realNL}$ was already defined in (3).

It can be easily observed that the total distortion error DFE_{total} defined in (4) is the sum of errors defined by (3) and (5).

Fig. 17 shows the mean and the CDF of the absolute value of the defined factor errors for the THD_V estimation in all the time intervals of a certain week. It can be observed that the total error is mainly influenced by the error of the estimation of the THD_V and that the error of the estimation of NL load categories has lower impact on the overall THD_V error. It can be also seen that during 90% of the time the total error is lower than 0.156 for feed-forward network and lower than 0.143 for autoregressive ANN.

Standard EN50160 [4] requires weekly 95th percentile of THD_V to comply with established limits. Fig. 18 shows the comparison between the 95th percentile value of real THD_V and THD_V predicted by the proposed method in ten simulated weeks, with a mean error of 9.8% and

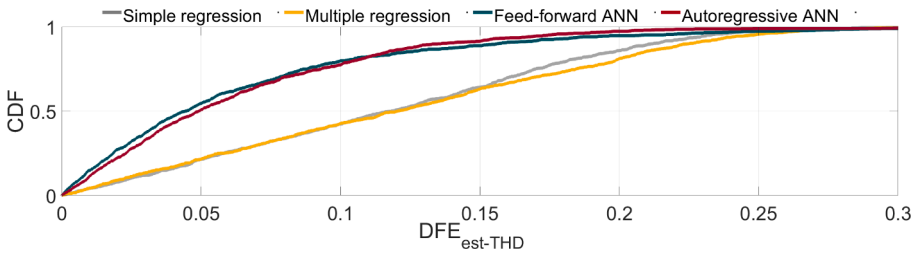

Fig. 12. Cumulative distribution of absolute errors in THD estimation ($DFE_{est-THD}$) with different algorithms at busbar TR.

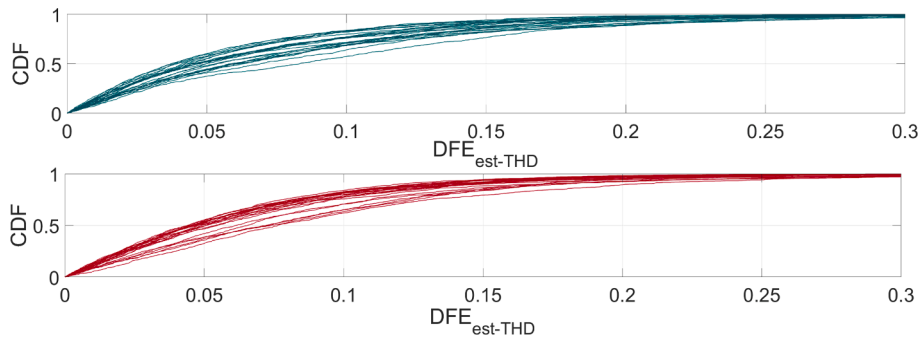

Fig. 13. Cumulative distribution of absolute errors in THD_V estimation ($DFE_{est-THD}$) in 24 weeks estimated by means of feed-forward ANN (top) and autoregressive ANN (bottom) at busbar TR.

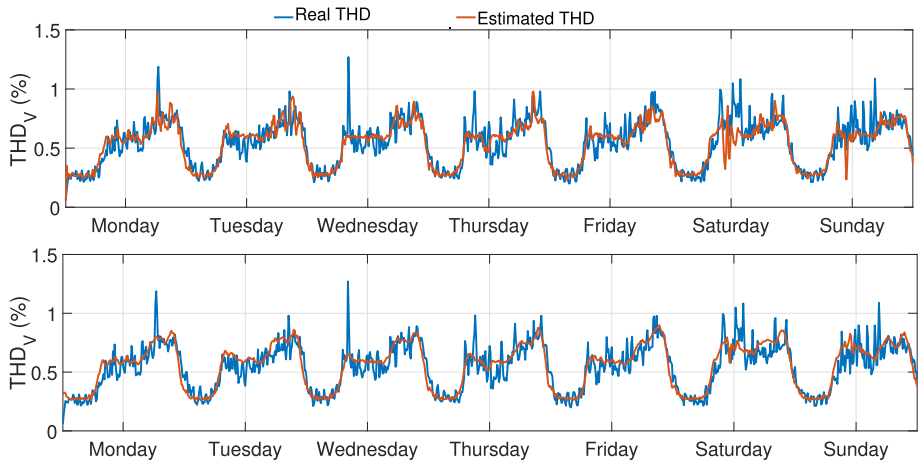

Fig. 14. Real and estimated THD_V by means of feed-forward ANN (top) and autoregressive (bottom) at secondary of transformer (bus TR).

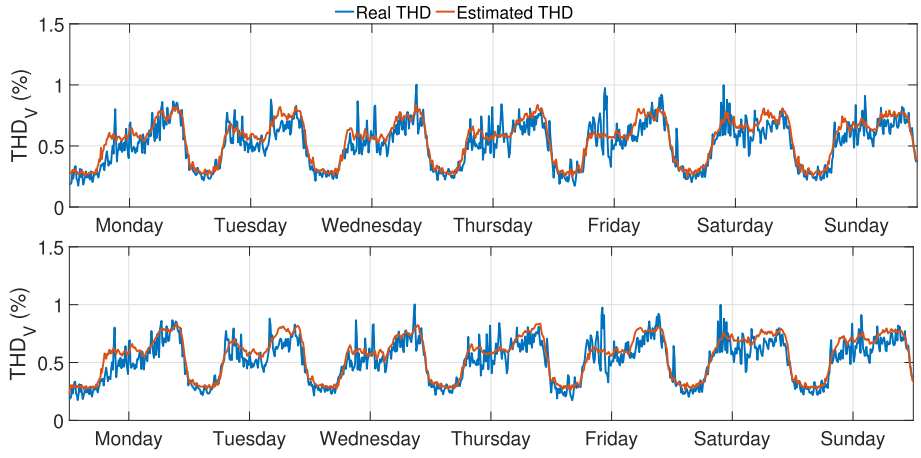

Fig. 15. Real and estimated THD_V by means of feed-forward ANN (top) and autoregressive (bottom) at busload 471 (bus LB).

standard deviation of 5.6 in estimations. As it can be seen, a close prediction is achieved with a slight underestimation which is caused by the existence of short THD_V peaks that are smoothed in the estimation, as shown in Fig. 14.

5.4. Application to scenario with higher levels of THD_V

THD_V can reach different levels depending on the penetration of nonlinear loads, short circuit capacity and operating condition of the network. In future scenarios with increased penetration of electric vehicles and other power electronic interfaced loads, storage and generation technologies as well as other nonlinear loads (e.g., efficient lighting), THD_V levels are expected to increase.

Results presented in previous sections referred to a scenario with low harmonic distortion. In this section, a new case study is presented for a scenario where higher voltage distortion levels are reached in order to verify the validity of the methodology in such situation. In this case study, the proposed estimation method is applied at bus 400 of the network, where THDV reaches values over 4% in the estimated week.

The estimation results are presented in Fig. 19, where the

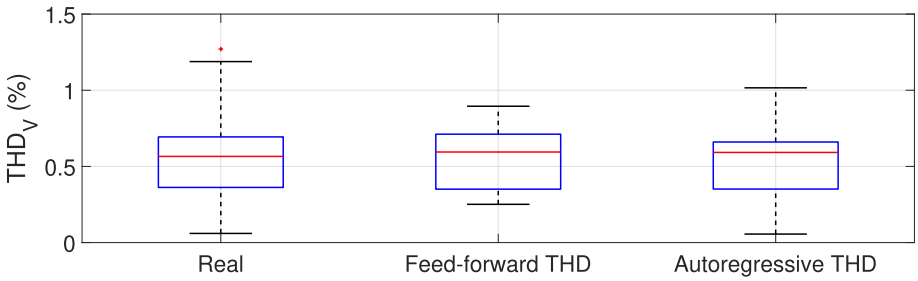

Fig. 16. Real and estimated THD_V boxplots at secondary of transformer (busbar TR).

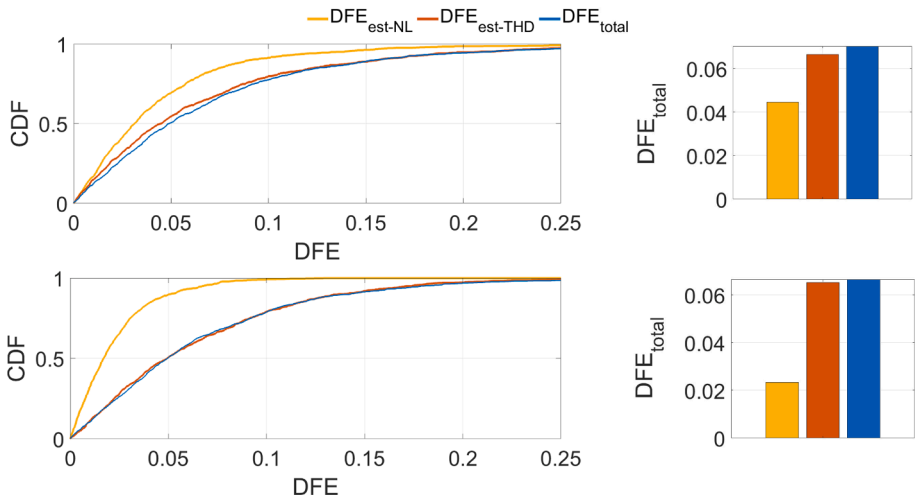

Fig. 17. CDF and mean of absolute factor errors errors during one week for THD estimated with feed-forward ANN (top) and autoregressive ANN (bottom) at secondary of transformer (busbar TR).

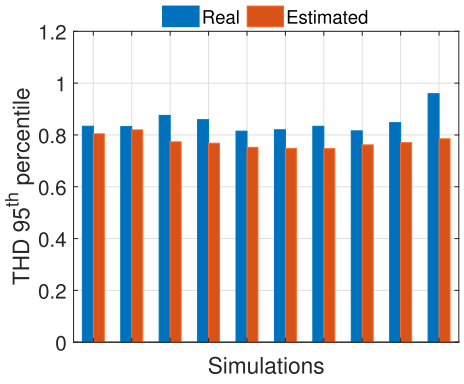

Fig. 18. THD 95th percentile values estimated with autoregressive ANN at secondary of transformer.

performance of both feed-forward and autoregressive ANNs is shown. As it can be seen in this figure, the method shows a similar accuracy to results previously presented.

Fig. 20 presents the boxplot of the THD_V values obtained for all 10-min time intervals of a week. It can be seen from Figs. 19 and 20 that both ANNs estimate the median, the 25th and the 75th percentile of THDV values with a good accuracy. The minimum value is also well estimated by both ANNs. The estimation of the highest THD_V values, occurring infrequently, is slightly less accurate though, as it was also observed in Fig. 16.

6. Conclusion

This paper presented a methodology to forecast THD_V at LV busbars in residential distribution networks. The proposed methodology uses limited number of smart meters with sub-metering functionality to predict demand disaggregation into linear and NL loads first, and then to estimate THD_V at different buses in the network. This methodology enables the integration of new functionalities into already installed monitoring devices in order to forecast present or future harmonic distortion.

Several algorithms have been tested and compared for both, decomposing demand and estimating THD_V. Autoregressive and feed-forward neural networks are shown to be the best and comparable methods to forecast THD_V considering both stages of estimation. It has been demonstrated that relatively small numbers of advanced smart meters in the network is enough for accurate harmonic estimations. Good performance is achieved with a 10% penetration of smart meters with sub metering functionality.

Other machine learning techniques, such as DNN or other ANN architectures, can be applied to solve the problem of harmonic forecasting. Trying different machine learning algorithms and comparing their performance with the approach used in this work has been identified as a potential future development. Another future area of work might be to explore possibility of development of a set generic trained ANNs for groups of typical networks and buses.

The methodology is illustrated here on a residential distribution network, but a similar process can be followed to predict THD_V in commercial, industrial or hybrid grids. It is important to note though that in order to predict load decomposition and THD_V in a particular network the ANNs should be trained in that network, but the parameters or model of the network are not required to be known.

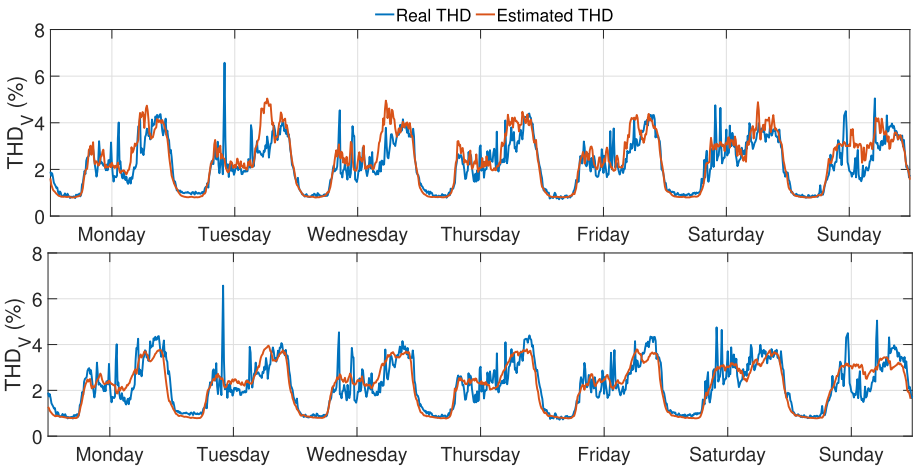

Fig. 19. Real and estimated THD_V by means of feed-forward ANN (top) and autoregressive (bottom) in highly distorted scenario.

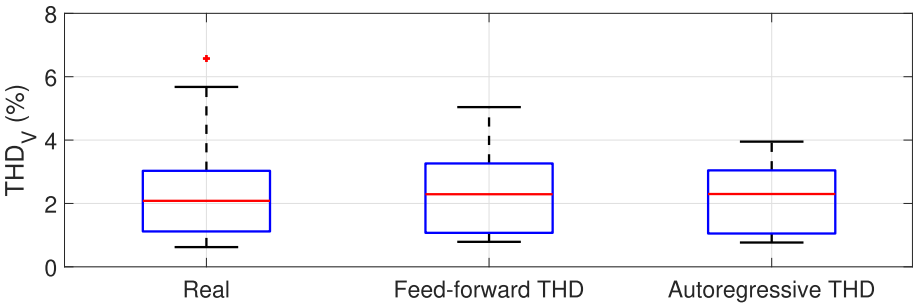

Fig. 20. Real and estimated THD_V boxplots in highly distorted scenario.

Finally, the approach is illustrated on a week-ahead prediction, but it can also be followed for real-time estimation. With this, PQ can be monitored on a continuous basis which allows the analysis of trends and the assessment of the compliance with the recommended limits.

7. Funding

This work was supported by the Ministerio de Ciencia, Innovación y Universidades, Spain, under Project RTI2018-097424-B-I00.

CRediT authorship contribution statement

Pablo Rodríguez-Pajarón: Conceptualization, Methodology, Validation, Software, Formal analysis, Writing - Original Draft. **Araceli Hernández:** Conceptualization, Methodology, Writing – original draft, Funding acquisition, Supervision. **Jovica V. Milanovic:** Conceptualization, Methodology, Writing – review & editing, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] U.S. Energy Information Administration, Annual energy outlook 2018 with projections to 2050, Washington, DC, USA; 2018. URL: https://www.eia.gov/outlooks/archive/aeo18/pdf/AEO2018.pdf.
[2] Aman M, Jasmon G, Mokhlis H, Bakar A. Analysis of the performance of domestic lighting lamps. Energy Policy 2013;52:482–500. Special Section: Transition Pathways to a Low Carbon Economy.
[3] Wang Y, Yong J, Sun Y, Xu W, Wong D. Characteristics of harmonic distortions in residential distribution systems. IEEE Trans Power Delivery 2017;32:1495–504.
[4] EN50160. Voltage Characteristics of Electricity Supplied by Public Distribution Networks, Standard EN 50160, CENELEC; 2010.
[5] IEC61000-3-2. Electromagnetic compatibility (EMC) - Part 3-2: Limits for harmonic current emissions (equipment input current ≤16 A per phase), Standard IEC 61000-3-2, CENELEC; 2018.
[6] Std P519. IEEE Recommended Practices and Requirements for Harmonic Control in Electrical Power Systems, Standard P519. IEEE; 2015. URL: https://ieeexplore.ieee.org/document/7050220.
[7] Salles D, Jiang C, Xu W, Freitas W, Mazin HE. Assessing the collective harmonic impact of modern residential loads—part I: Methodology. IEEE Trans Power Delivery 2012;27:1937–46.
[8] Rodríguez-Pajarón P, Hernández A, Milanovic JV. Probabilistic assessment of the impact of electric vehicles and nonlinear loads on power quality in residential networks. Int J Electr Power Energy Syst 2021;129:106807.
[9] Ye GG, Nijhuis MM, Cuk VV, Cobben JS. Stochastic residential harmonic source modeling for grid impact studies. Energies 2017;10:372.
[10] Wang Y, O'Connell RM, Brownfield G. Modeling and prediction of distribution system voltage distortion caused by nonlinear residential loads. IEEE Trans Power Delivery 2001;16:744–51.
[11] Au MT, Milanović JV. Establishing harmonic distortion level of distribution network based on stochastic aggregate harmonic load models. IEEE Trans Power Delivery 2007;22:1086–92.
[12] Au MT, Milanović JV. Development of stochastic aggregate harmonic load model based on field measurements. IEEE Trans Power Delivery 2007;22:323–30.
[13] Silva MM, Gonzalez MLY, Uturbey W, Carrano EG, Silva SR. Evaluating harmonic voltage distortion in load-variating unbalanced networks using monte carlo simulations. Transmiss Distrib IET Generat 2015;9:855–65.
[14] Zhou W, Ardakanian O, Zhang H, Yuan Y. Bayesian learning-based harmonic state estimation in distribution systems with smart meter and dpmu data. IEEE Trans Smart Grid 2020;11:832–45.
[15] Melo ID, Pereira JL, Variz AM, Garcia PA. Harmonic state estimation for distribution networks using phasor measurement units. Electr Power Syst Res 2017;147:133–44.
[16] Melo ID, Pereira JL, Ribeiro PF, Variz AM, Oliveira BC. Harmonic state estimation for distribution systems based on optimization models considering daily load profiles. Electr Power Syst Res 2019;170:303–16.
[17] Albadi M, El-Saadany E. A summary of demand response in electricity markets. Electr Power Syst Res 2008;78:1989–96.
[18] Ponoćko J, Milanović JV. Forecasting demand flexibility of aggregated residential load using smart meter data. IEEE Trans Power Syst 2018;33:5446–55.
[19] Ponocko J, Milanovic JV. Data requirements for a reliable demand decomposition in sparsely monitored power networks. In: 2018 IEEE PES Innovative Smart Grid

Technologies Conference Europe (ISGT-Europe); 2018. p. 1–6. doi: 10.1109/ISGTEurope.2018.8571585.
[20] Xu Y, Milanović JV. Day-ahead prediction and shaping of dynamic response of demand at bulk supply points. IEEE Trans Power Syst 2016;31:3100–8.
[21] Xu Y, Milanović JV. Artificial-intelligence-based methodology for load disaggregation at bulk supply point. IEEE Trans Power Syst 2015;30:795–803.
[22] Pipattanasomporn M, Kuzlu M, Rahman S, Teklu Y. Load profiles of selected major household appliances and their demand response opportunities. IEEE Trans Smart Grid 2014;5:742–50.
[23] Kolter JZ, Johnson MJ. REDD: A public data set for energy disaggregation research. In: Proc. Workshop Data Mining Appl. Sustain., San Diego, CA, USA; 2011. p. 59–62.
[24] Pecan Street Inc., Dataport; 2017. URL: http://www.pecanstreet.org/.
[25] Sedhom BE, El-Saadawi MM, El Moursi M, Hassan M, Eladl AA. Iot-based optimal demand side management and control scheme for smart microgrid. Int J Electr Power Energy Syst 2021;127:106674.
[26] Kuzlu M, Pipattanasomporn M, Rahman S. Hardware demonstration of a home energy management system for demand response applications. IEEE Trans Smart Grid 2012;3:1704–11.
[27] Wang X, Mao X, Khodaei H. A multi-objective home energy management system based on internet of things and optimization algorithms. J Build Eng 2021;33:101603.
[28] Srinivasan D, Ng W, Liew A. Neural-network-based signature recognition for harmonic source identification. IEEE Trans Power Delivery 2006;21:398–405.
[29] Rahimpour A, Qi H, Fugate D, Kuruganti T. Non-intrusive energy disaggregation using non-negative matrix factorization with sum-to-k constraint. IEEE Trans Power Syst 2017;32:4430–41.
[30] Kong W, Dong ZY, Ma J, Hill DJ, Zhao J, Luo F. An extensible approach for non-intrusive load disaggregation with smart meter data. IEEE Trans Smart Grid 2018;9:3362–72.
[31] Singh S, Majumdar A. Deep sparse coding for non–intrusive load monitoring. IEEE Trans Smart Grid 2018;9:4669–78.
[32] Capasso A, Grattieri W, Lamedica R, Prudenzi A. A bottom-up approach to residential load modeling. IEEE Trans Power Syst 1994;9:957–64.
[33] Cavallini A, Montanari GC, Cacciari M. Stochastic evaluation of harmonics at network buses. IEEE Trans Power Delivery 1995;10:1606–13.
[34] Blanco Castañeda A. Stochastic Harmonic Emission Model of Aggregate Residential Customers [Ph.D. thesis]. Fakultät Elektrotechnik und Informationstechnik; 2017.
[35] Collin AJ, Hernando-Gil I, Acosta JL, Djokic SZ. An 11 kv steady state residential aggregate load model. part 1: Aggregation methodology. In: 2011 IEEE Trondheim PowerTech; 2011. p. 1–8.
[36] Beale MH, Hagan MT, Demuth HB. Matlab Deep Learning Toolbox. Natick, MA, USA; The MathWorks, Inc.; 2017.
[37] Richardson I, Thomson M, Infield D, Clifford C. Domestic electricity use: A high-resolution energy demand model. Energy Build 2010;42:1878–87.
[38] Blanco AM, Gasch E, Meyer J, Schegner P. Web-based platform for exchanging harmonic emission measurements of electronic equipment. In: 2012 IEEE 15th ICHQP; 2012. p. 943–8.
[39] Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer series in statistics. Springer; 2009. URL: https://web.stanford.edu/hastie/ElemStatLearn/.
[40] Distribution System Analysis Subcommitteee: Distribution Test Feeder Working Group. IEEE european low voltage test feeder, webpage; 2015. Available: http://sites.ieee.org/pes-testfeeders/resources/.
[41] Dugan RC, McDermott TE. An open source platform for collaborating on smart grid research. In: 2011 IEEE PES General Meeting; 2011. p. 1–5. doi: 10.1109/PES.2011.6039829.
[42] D3.4 Smart meters architecture and data model analysis. NOBEL GRID project; 2016. URL: http://nobelgrid.eu/deliverables/.
[43] Bosovic A, Renner H, Abart A, Traxler E, Meyer J, Domagk M, et al. Deterministic aggregated harmonic source models for harmonic analysis of large medium voltage distribution networks. IET Gener Transmiss Distrib 2019;13:4421–30.
[44] Xu Y, Milanović JV. Accuracy of ann based methodology forload composition forecasting at bulk supply buses. In: Proc. Int. Conf. Probab. Methods Appl. Power Syst.; 2014. p. 1–6.
[45] Skansi S. Introduction to Deep Learning From Logical Calculus to Artificial Intelligence, Undergraduate Topics in Computer Science, 1st ed.; 2018.
[46] Xu Y. Probabilistic Estimation and Prediction of the Dynamic Response of the Demand at Bulk Supply Points [Ph.D. thesis]. Manchester, U.K.: School Elect. Electron. Eng., Univ. Manchester; 2015.
[47] Buitrago J, Asfour S. Short-term forecasting of electric loads using nonlinear autoregressive artificial neural networks with exogenous vector inputs. Energies 2017;10.
[48] Haykin S. Neural networks: a comprehensive foundation. 1999. New Jersey: Mc Millan; 2010. p. 1–24.
[49] Heumann C, Schomaker Shalabh M. Introduction to Statistics and Data Analysis. Cham, Switzerland: Springer; 2016. p. 249–90.
[50] Instituto Nacional de Estadística, España, Proyección de hogares 2018; 2018. Available: http://www.ine.es/dyngs/INEbase/listaoperaciones.htm.
[51] Rigoni V, Ochoa LF, Chicco G, Navarro-Espinosa A, Gozel T. Representative residential LV feeders: A case study for the north west of England. IEEE Trans Power Syst 2016;31:348–60.
[52] Koirala A, Suárez-Ramón L, Mohamed B, Arboleya P. Non-synthetic european low voltage test system. Int J Electr Power Energy Syst 2020;118:105712.
[53] Dugan RC, Montenegro D. Reference guide. The open distribution system simulator (OpenDSS). Electrical Power Research Institute, Inc.; 2018.