Multiple Households Very Short-Term Load Forecasting using Bayesian Networks

Michel Bessani
Department of Electrical Engineering
Federal University of Minas Gerais, UFMG
Belo Horizonte, Brazil
mbessani@ufmg.br

Julio A.D. Massignan, Talysson M. O. Santos
João B.A. London Jr., Carlos D. Maciel
Department of Electrical and Computing Engineering
University of São Paulo, USP
São Carlos, Brazil
{julio.massignan, talyssonsantos}@usp.br
{jbalj, maciel}@sc.usp.br

***Abstract*—Load forecasting is essential for different activities on power systems, and there is extensive research on approaches for forecasting in different time horizons, from next-hour to decades. However, because of higher uncertainty and variability compared to aggregated or medium and high voltage, the forecasting of the individual household load is a current challenge. This paper presents a load forecasting for multiple households using Bayesian networks. Our model, which is multivariate, uses past consumption, temperature, socioeconomic and electricity usage aspects to forecast the next instant household load value. It was tested using real data from the Irish smart meter project and its performance was compared with other forecasting methods. Results have shown that the proposed approach provides consistent single forecast model for hundreds of households with different consumption patterns, showing a generalisation capability in an efficient manner.**

***Index Terms*—Bayesian Networks; Data-driven Modelling; Household Load Forecasting; Very Short-term Load Forecasting; Smart Meters.**

I. Introduction

Load forecasting (LF) is useful for multiple activities on power systems, by providing the expected demand it can assist scheduling, control, operation, maintenance activities, and planning for the generation, transmission and distribution systems [1], [2], [3], [4]. The LF task can be divided according to the prediction horizon [1], [4]: Long-Term Load Forecasting (LTLF) ranging from years to decades; Medium-Term Load Forecasting (MTLF) ranging from months to few years; Short-Term Load Forecasting (STLF) ranging from days to weeks; and Very Short-Term Load Forecasting (VSTLF) ranging from minutes to the next hour.

Besides the relevant research in this topic, there still are challenges for LF [5], in specific for individual households LF, which present an intrinsic higher uncertainty and volatility [6] compared to the aggregated, medium and high voltage networks LF. In addition, there is also a large amount of data arising from smart meters at households together with the multivariate data related to anthropological and structural information [7]. Besides such aspects, there also is the need for generalist models capable of performing multiple households forecast instead of specific ones for each building LF [8]. Households LF is also motivated by other modern power systems applications, as demand response [8], that can benefit from a real-time LF.

There are different approaches that resulted in small error values for household load forecasting, which are evaluated with different forecasting metrics [8], [9], [5], [10]. However, these methods have some aspects that must be highlighted. The use of invasive information related to household inhabitants routine and also the households' explicit identification to perform forecasting. Forecasting methodologies that need all the previous observed load values to the forecasting model learning. And the models that are optimised to individual building load forecasting. These aspects are related to intrusive information usage, or a large amount of previous load consumption data, or are specific learned to individual household forecasting.

In this sense, this paper presents the use of Bayesian Networks, a probabilistic graphical model [11], to perform multiple household LF by using multivariate data related to non-invasive households socioeconomic and electricity usage aspects, the temperature, and their previous power consumption measures. The model was evaluated using the Irish smart meter project data [12] by forecasting the next half-hour consumption for hundreds of households in all instants for different days. The proposed approach comprises a training stage, where model learning is performed by using the past two weeks load data for all households together with some socioeconomic and electrical usage features. The forecasting stage tests how such the model predicts the next demand value for each household given the previous consumption and without using the households identification.

The rest of this manuscript is organised as follows. Section II reviews the related research, Bayesian Networks are

This work was supported by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, Fundação de Amparo à Pesquisa do Estado de São Paulo - Brasil (FAPESP) - grants no. 2014/50851-0, 2016/19646-6, and 2018/00214-4, and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) - Projects 465755/2014-3 and 308297/2018-0. The authors would like to acknowledge the CER Smart Metering Project - Electricity Customer Behaviour Trial, 2009-2010.

formally introduced in Section III, Section IV presents the materials and methods, as the smart meter data, followed by the data pre-processing, the load forecasting trough Bayesian Networks, and how this approach forecasting capability was evaluated. Section V presents and discusses the results of our model considering the gaps mentioned previously, and Section VI concludes this manuscript.

II. Literature Review

Because of the importance of a reliable prediction for the operation and planning of energy systems, there is extensive research on LF [4], [6], [13]. In specific, the household VSTLF, addressed in this manuscript, is useful for demand-side management [8], demand response by a real-time stimulus for handling unexpected changes [14], [1], and also to assist system restoration during contingencies [15]. The approaches for LF range from linear, like linear regression and auto-regressive models, to non-linear, as artificial neural networks (ANN) and support vector machines, and are applied in the high, medium and low-voltage systems. The LF for low-voltage systems can also be divided into aggregated load or individual household load forecasting, this last being the focus of this paper.

Chaouch [9] performed a household STLF by using a combination of clustering and functional wavelet-kernel approach to predict the daily load curve of each cluster. They use all the previous observed power consumption as learning set to forecast the next day load curve and also evaluated the performance with the Irish smart meter data. Javed et al. [8] used a multilayer perceptron to perform household STLF by combining house inhabitants behaviour and residences features, which resulted in an efficient forecast for multiple houses. Another study that presented a model for a single household was presented in [16], which used an ANN together with individual appliances information, resulting in good performance results. Such approaches that depend on appliances information or detailed inhabitants and houses features should be carefully addressed since privacy issues are an essential aspect of modern power systems [17].

A trend in non-linear models is the use of Deep Learning (DL), as in [5], where a DL approach to STLF of household consumers was presented using the Irish smart meter project data. This approach was compared to other models, and the achieved results showed that the proposed DL overcomes the others. Besides the small errors reported, their model was tested in only a specific month, without assessing the effects of seasonality, an essential factor in LF. Almalaq and Zhang [10] presented a DL forecasting model, which parameters were optimised by an evolutionary heuristic, providing excellent results. However, only a single residence and a single commercial building were used separately to evaluate their approach, which limits the generalisation of the reported results.

Bassamzadeh and Ghanem [18] used a Bayesian Network to perform load modelling for aggregated and residential load. They randomly divided the data into two sets, 90% of data for the model training, and the remaining for testing, without using the consumption' temporal sorting during prediction tests. As Bayesian networks can model the relationship among random variables, their study evaluated the effect of different pricing policies on electrical consumption using a data set with the consumption of 25 customers. Although [18] study does not perform household load forecasting, it is the only one that used a Bayesian Network to model residential power consumption at the individual household level.

There are also other LF studies that used Bayesian methods in different approaches [19], [20], [21], [22]. In [19], a Bayesian combined predictor is used to decide which model will provide the forecast. In [21] and [22], the Bayesian method is used as a probabilistic approach to minimise the error during ANN learning, and [20] used a point Bayesian estimation to forecasting the day-ahead peak loads. All these models forecast a single load value related to the aggregated load from large regions, i.e., with less volatility than individual households load.

III. Bayesian Networks

A Bayesian network (BN) is a probabilistic graphical model [11] defined as a directed acyclic graph (DAG), $\mathcal{G} = (\mathbf{V}, \mathbf{E})$, where $\mathbf{V}$ is a set of vertices representing n random variables $\mathbf{X} = \{X_1, ..., X_n\}$, and $\mathbf{E}$ is a set of m directed edges representing causal and influential relationship among the variables together with the joint probability $P(X_1, ..., X_n)$. Besides $\mathcal{G}$, the BN model also has a set of conditional probabilities distribution between the dependent (connected) variables - $\Theta = \{\theta_1, ...\theta_n\}$. According to the local Markov property [23], each vertex is conditionally independent of its non-descendants given its parents states. As a consequence, a vertex state can be represented by a conditional probability, $\theta_i = P(X_i \mid Pa(X_i))$, where $Pa(X_i)$ are the parents vertices of X_i. Using this, we can rewrite the joint probability using the chain rule as [24]:

$$P(X_i, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid Pa(X_i)), \tag{1}$$

in this manner, the joint probability can be calculated by using the set of conditional probability distributions - Θ.

A BN model directly depends on the graph $\mathcal{G}$ and the conditional probabilities Θ, and both must be learned. The graph learning, also known as structural learning, can be Constraint-based, Score-based and hybrid [25]. Constraint-based uses independence tests to evaluate edges in $\mathcal{G}$. Score-based use heuristics to search for structures that are evaluated using some goodness-of-fit metric as Akaike Information Criteria (AIC) or Bayesian Information Criteria (BIC), and hybrid methods combine these two approaches. Learning strategy examples are super-structures learning [26], model averaging [27], hill climbing [28], and tabu search [11].

After learning the $\mathcal{G}$, it is also necessary to learn the quantitative part of the BN, Θ, that depends on the edges in $\mathcal{G}$ and also on the data being modelled, which can be discrete, continuous, or a combination of both. The choice of these types affects the conditional probabilities learning strategy. For discrete data is necessary to learn the conditional probability tables (CPTs), where each one describes the probabilities of each state of a variable given its parents' states, resulting in a non-parametric model. For continuous ones, some class of probability distributions (e.g., Gaussian) must be assumed, and their parameters must also be estimated. This learning procedure can be performed by maximum likelihood or also a Bayesian estimation [11]

With a complete BN model learned, it is possible to use it to compute the conditional probability distribution of an unobserved variable given some evidence - $P(X_i \mid \mathbf{e})$, where $\mathbf{e}$ are the observed variables states. The unobserved variable can also be estimated, which is the value that maximises the *posteriori* probability distribution:

$$\hat{x}_i = \arg\max_{x_i} P(X_i \mid \mathbf{e}). \quad (2)$$

IV. Materials and Methods

A. Smart Meter Data

The Irish smart meter data used in this study is publicly available and is the result of the smart metering electricity customer behaviour trials performed by the Irish Commission for Energy Regulation (CER) [12]. The full data set includes over 5,000 Irish homes and small and medium enterprises, and comprises to active power consumption records from 1st July 2009 to 31st December 2010 with a 30 minutes sampling rate, together with questionnaires and survey responses about households occupants and electricity usage aspects features, and individual customers' classification with type (residential or enterprise), tariff scheme, and tariff incentive stimulus.

Besides the power consumption time series, the following trials survey answers related to electricity usage and social aspects were selected: number of inhabitants, social class based on the occupation of chief income in the household, and electricity use for cooking and house heating. These were chosen considering the findings presented in [29], where a factorial analysis of the variables related to social class and electricity usage aspects was carried out to model their relationship with energy consumption. An important aspect is that information about specific appliances or inhabitants' habits was not used. The respective temperature for the instant of power consumption measures, which was showed as a relevant feature for household consumption LF by [8], was also used and was obtained through the Irish Meteorological Service [30].

B. Data pre-processing

The power consumption data from CER smart meter project [12] was first filtered to represent only the meters for households that belong to the tariff and stimulus control group, and that answered the survey, resulting in 929 metered households. After that, 59 evenly spaced dates were chosen, starting from 2009-08-01 and ending at 2010-11-21. For each one of these dates two data set were formed, one to learn the VSTLF model consisting of two weeks of measurements starting at 00:00:00 in the dates previously mentioned, and the other composed of the next day measurements to test the model. This process resulted in 59 pairs of learning and test sets. In this step of generating the learn and test sets, the smart meters that present missing values were removed.

First, the power consumption values at each train and test sets were normalised by dividing the consumption by the respective household average consumption observed in the train data set. Then, similar to [18], the normalised consumption was then discretised. The use of discretised values was motivated by the application of discrete BNs, which are a non-parametric model that can be adequate to different data behavior without modifications. This aspect was already highlighted in [18], where a comparison between discrete and Gaussian BN for modelling households' electricity consumption indicated equivalent results. Here, we use a quantile-based discretisation due to its capability to preserve the intervals meaning independently of data drift [31], with 100 intervals calculated from the train set. These intervals were also used to discretise the normalised test set.

With the discrete consumption data, the temporal associations in the consumption time series were investigated by using mutual information (MI). This was done similar to other studies [32], [33], [34], [35], and the MI measures the amount of information that the original time series contains about its lagged version calculated as:

$$I_k\left(C_t; C_{t-k}\right) = \sum_{x\in\Re}\sum_{y\in\Re} p\left(x,y\right) \log\frac{p\left(x,y\right)}{p\left(x\right)p\left(y\right)} \quad (3)$$

where C_t is the consumption at time instant t, C_{t-k} is the consumption k samples before t, $p(x,y)$ is the joint probability mass function of C_t and C_{t-k}, $p(x)$ is the marginal probability mass function of C_t and $p(y)$ is the marginal probability mass function of C_{t-k}. The higher the MI (I_k) value, more information can be obtained about C_t from C_{t-k}, i.e, due to the knowledge of C_{t-k} the uncertainty of C_t reduces [36]. The mutual information was used to choose how many lags (past power consumption values) should be included in the learning set that then is used to fit the BN model during the learning process described in the following Subsection.

The temperature values were also discretised, but using 30 equal size intervals. These intervals were calculated using the temperatures observed in the train set period and were also used to discretise the temperature for the respective test set. The discretised power consumption and past consumptions with higher mutual information, temperature, time index, and the survey responses compose the data used to learn the BN models for each train set.

C. Load Forecasting trough Bayesian Networks

The proposed VSTLF consists of an initial training stage, where learning is performed by using the pre-processed past information gathered in each one of the learning sets for all households. This process results in an unique BN model $\mathrm{BN}_C = (\mathcal{G}_C, \Theta_C)$ that relates the discretised load demand at each time sample (C_t) with the remind of variables (vertices) in $\mathcal{G}_{\mathcal{C}}$ for each learning set:

- C_{t-k}: the discretised consumption k samples before t that presented the higher MI (I_k) values;
- T_{ind}: the time index of each sample, ranging from 1 to 48 representing half hour steps during a day;
- DT: the current day type (weekday or weekend);
- $Temp$: the discretised temperature values;
- Features from Survey ($\mathbf{F-Survey}$): the number of inhabitants (NI), the social class (SC), electricity use for house heating (HH) and cooking (EC).

The edges in $\mathbf{E}$, that represent the relationship among these variables, were obtained by a score-based structure learning method. The tabu search heuristic, which is capable of escaping from local optimums, was used together with the AIC score, which tends to result in models with a good predictive performance [37].

With the $\mathcal{G}_{\mathcal{C}}$ learned, the Θ_C can be obtained from the same data by respecting the dependencies described in $\mathbf{E}$. As the consumption data was discretised, each $\theta_i \in \Theta_C$ will be a CPT describing the probabilities for each X_i state conditioned to its parents' states. In our approach, Bayesian estimation of θ_i was used, where a prior distribution π_i is combined with the relative frequencies observed on data. This π_i was assumed a uniform distribution and can be interpreted as an imaginary sample [24], [38] that is combined with observed data during the relative frequencies calculation $\hat{P}(X, Pa(X_i))$ and $\hat{P}(Pa(X_i))$, which are used to calculate the CPTs in Θ_C.

Following the learning of BN_C, the inference procedure, i.e., the forecasting itself, was performed by averaging likelihood weighting [11], which consists of weighting the evidence by their likelihood during a forward sampling of BN_C. The sampling starts on the variables without parents and forward until sampling from the ones without children, and then the simulated observations are used to estimate $\hat{P}(C_t^i \mid e)$. The forecasting stage uses the learned BN together with the latest sampled information to estimate the next demand value C_t^i for the *i-th* household given the latest sampled values $\{C_{t-k}^i, Temp\}_{latest}$ and the known features $\{T_{ind}, DT, \mathbf{F-Survey^i}\}$. After that, the discrete consumption forecast at t for the *i-th* household ($\hat{C}_t^i$) is the value that maximises the estimated *posteriori*.

In Fig. 1, a summary of the steps described above in this section are presented as a Flowchart divided in three parts, Data pre-processing, BN structural and parameters learning, and the load forecasting stage.

D. Performance Evaluation

The evaluation of the VSTLF model was performed by metrics already used in other studies that addressed multiple or individual household load forecasting [8], [18], [5], [9], [10]: The Normalized Root Mean Squared Error (NRMSE), which balances penalization of large forecast errors and forecast data variability showing how well the model is forecasting the real mean; The Mean Absolute Error (MAE) and the Median Absolute Error (MedAE), this last being robust to outliers,

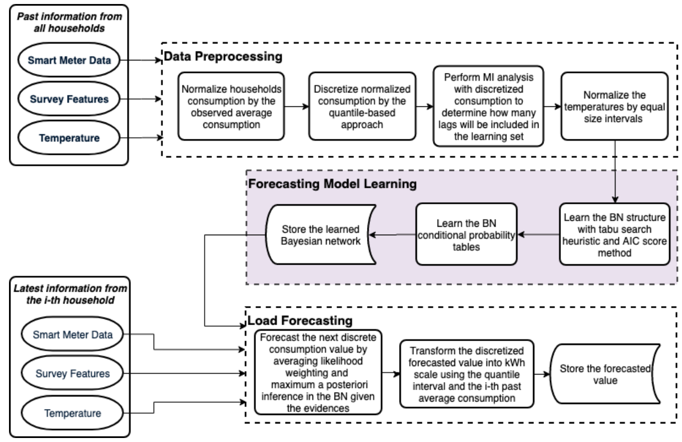

Fig. 1. Flowchart of the proposed method for load forecasting of households using the presented BN approach.

and both providing a value in the same scale of the forecasting quantity, in this case, kWh; and the Mean Arctangent Absolute Percentage Error (MAAPE) that measures the slope of the hypotenuse formed by the absolute forecast error and the real value [39]. These are computed as follows:

$$\text{NRMSE} = \sqrt{\frac{\frac{1}{N}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{y_{\max} - y_{\min}}}, \tag{4}$$

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N} \mid y_i - \hat{y}_i \mid, \tag{5}$$

$$\text{MedAE} = \text{median}\left(\mid y_1 - \hat{y}_1 \mid, \ldots, \mid y_n - \hat{y}_n \mid\right), \tag{6}$$

$$\text{MAAPE} = \frac{1}{N}\sum_{i=1}^{N} \arctan\left(\frac{\mid y_i - \hat{y}_i \mid}{y_i}\right), \tag{7}$$

where $\hat{y}_i$ is the forecast value, y_i is the real value, N is the number of forecasts realised, $y_{\max}$ and $y_{\min}$ are the maximum and minimum values observed in the test set.

When evaluating household load forecasting, the use of the widely recognized evaluation metric Mean Absolute Percentage Error (MAPE) faces some issues that are related to an intermittent aspect because of low, or almost zero, load values during different instants [7], resulting in infinite or very large numbers inside the summation. This issue was the motivation of using the MAAPE [39], an alternative metric to the absolute percentage error but from a different perspective that can handle the intermittent behaviour of households load.

In addition to performance evaluation by such metrics, the proposed BN was also evaluated by comparison with the following methods: Persistence technique [40], which assumes the next value will be equal to the previous one observed; Multilayer Perceptron (MLP) that is a class of feedforward ANN method [8], [41] implemented using the scikit-learn python package [42]; and Hidden Markov Model (HMM) with Gaussian emissions [43], [44] implemented using the hmmlearn python package. All these models were learned and evaluated in the same manner as the BN model, this learning step is the grey part of the flowchart presented in Fig. 1.

V. Results

MI analysis highlighted a pattern present in the households consumption time series. This was used to define the past consumption features C_{t-k} that composes the data used during the BN models learning and LF. This analysis indicated that values at a half-hour before (C_{t-1}), one hour before (C_{t-2}), and an hour and a half before (C_{t-3}) together with the same consumption in the same instant at the past day (C_{t-48}) are the ones with higher MI. Important to mention that this pattern already had been reported for other power consumption time series at different aggregation levels [45], [46], [47]. Our approach does not use the MI during BN model learning, but the past power consumption values with higher MI values compose the learning and test sets.

Fig. 2 illustrated the MI of three households' power consumption time series for different lag values. The first lag (C_{t-1}) is the one with the higher mutual information. The MI decreases until around the lag 24 (12 hours before), and then increases until a peak at lag 48 (24 hours before), and after the lag 48, this behaviour is periodic with cycles of 48 lags. The MI analysis also highlighted that exists variability in how consumption is related to past consumption measures for the households, reflecting the diversity of power consumption profile present in the data set.

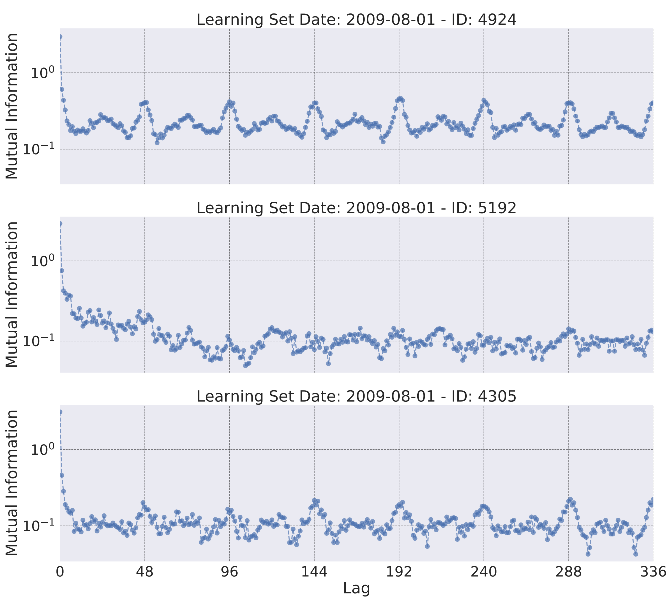

Fig. 2. Mutual information for the discrete power consumption time series of single households at different lag values. The scale on the y-axis is logarithmic.

Following this, these past values ($C_{t-1}, C_{t-2}, C_{t-3}, C_{t-48}$) composes the final data set together with the consumption C_t at the time index $\mathrm{T_{ind}}$ and, DT, Temp, together with the survey responses NI, SC, HH, and EC. The Fig. 3 illustrates the DAGs $\mathcal{G}_C$ obtained for two learned sets, where is possible to note the relationship between these variables in both graphs, as the survey responses and the electricity usage, the consumption at a time index and past consumption values, and the indirect association between consumption and temperature through the time index. Besides some differences, these relationships were consistent with the structures learned from the different training sets.

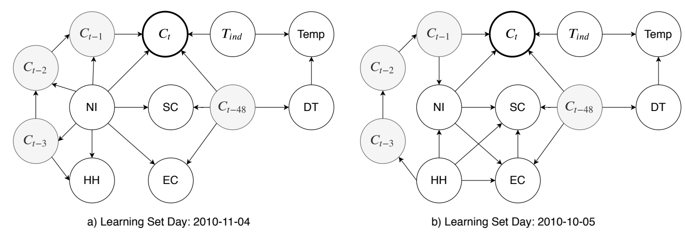

Fig. 3. BN structures obtained for two different train sets. Besides the difference in some edges, the structures highlights a persistent relationship among the used variables. The gray nodes are related to past consumption information, and the C_t node is highlighted with a bold border.

With the BN models fitted for each learning set and the necessary evidence $\{C^i_{t-k}, Temp, T_{ind}, DT, F-Survey^i\}$ present in the test set, the LF of the next C^i_t was carried out. This same approach of model learning was performed for the persistence, ANN, and HMM forecasting methods. Before presenting the results and evaluation metrics for these four methods using the complete test set, an illustration of two forecasts performed by them is presented in Fig. 4.

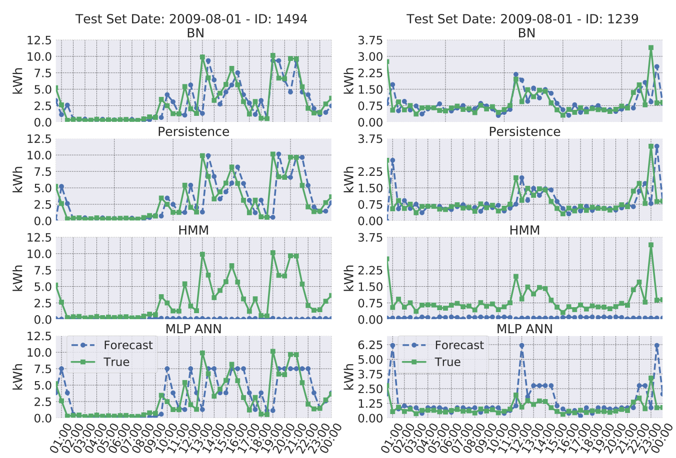

Fig. 4. Illustration of load forecasting for the proposed BN method and the persistence, HMM, and ANN for two households with different consumption patterns at one test set.

As the plots highlight, the HMM forecasts mostly a consumption equals to zero, which is the value with a higher probability of occurrence in the power consumption time series. The BN, persistence and ANN methods resulted in better forecasts, being the ANN the one with the worst result among them, while the BN forecasting capability is better than the persistence one. The persistence provides good results as a consequence of the slow variability of the households power consumption and the presence of intervals with constant zero, or near-zero, consumption. However the BN model showed better accuracy in handling such slow variability, including the different peak demands, as in the evening for ID:1494, and in the lunchtime for the ID:1239, and also the low values. This reflects the generalization capability of the proposed BN model, which was learned using the data of all the households and without knowing their ID.

In Fig. 5, a box plot of the observed errors in kWh for forecasting of the 48 next power consumption in each test set and for each forecasting method is presented, which resulted in 59 evaluations for each method. The errors for the ANN resulted in the highest values, followed by the HMM. We can also observe that the BN and persistence methods resulted in the lowest absolute median errors (smaller than 0.5 kWh).

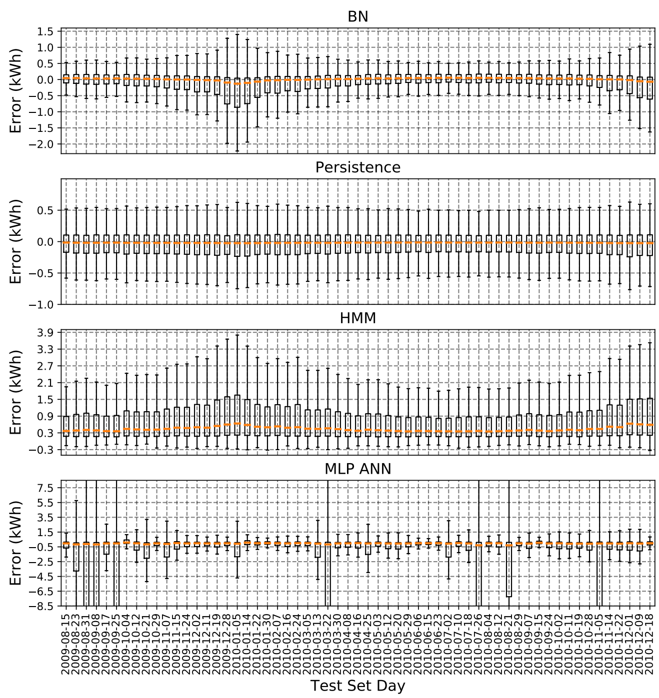

Fig. 5. Boxplot of the observed errors during the forecast for the 59 test sets using the proposed BN, and persistence, HMM and MLP ANN.

An exciting aspect present in these forecasting errors visualisation is that the error increases for dates near to the year-end and beginning for the BN, persistence and HMM. This phenomenon was also present in [9] results for this same data set. A hypothesis about such increased error is the occurrence of extremes winter on 2009-10 [48] and 2010-2011 [49] in Ireland. These extreme weather events naturally contributed to an unusual electricity consumption increase in households' electricity usage during these periods. This is a consequence of more time inside the houses using more energy, which includes cooking, and greater energy demand for house and water heating.

In Table I, a summary of the values observed for the forecasting performance metrics NRMSE, MAE, MedAE and MAAPE is presented for the BN and the other three forecasting methods used for comparison. The bold values indicate which method resulted in the lowest median and average value for each performance metric. The ANN model did not result in any of the lowest values observed for all the metrics calculated. Another important point is that the HMM, which forecasts the load as zeros due to its higher probability of occurrence, resulted in the smallest MAE median and average values. The persistence method resulted in the smallest MedAE and MAAPE values, being the observed values of such metrics for the BN slightly higher, and the BN presented the smallest NRMSE values.

By evaluating these metrics and the Figs. 4 and 5, it is evident that the choice of a metric to evaluate the forecasting of such high intermittent time series may result in erroneous interpretations, as the better performance of the HMM indicated by the MAE. Another point is the good performance of the persistence method indicated by the MedAE and MAAPE, with values slight better than the ones observed for the BN method. This is directly related to the fact that these two metrics are robust to extreme errors, which naturally happens for persistence forecasting when there are fast changes in the time series [40]. The NRMSE, which penalises these extreme errors, indicated that the proposed BN resulted in the better forecasting.

TABLE I
PERFORMANCE METRICS CALCULATED USING THE KWH FORECAST OF HOUSEHOLDS' CONSUMPTION FOR THE 59 TEST SETS USED FOR VSTLF.

Bayesian network				
Error Metric	NRMSE	MAE(kWh)	MedAE (kWh)	MAAPE
Median	**0.0877**	1.0085	0.1540	0.5035
Average	**0.1962**	1.6500	0.1667	0.5001
Std	0.2615	1.5550	0.0338	0.0378
Persistence				
Error Metric	NRMSE	MAE (kWh)	MedAE (kWh)	MAAPE
Median	1.1037	1.0931	**0.1450**	**0.4395**
Average	2.4453	1.901	**0.1457**	**0.4377**
Std	3.1605	1.9741	0.0101	0.0143
HMM				
Error Metric	NRMSE	MAE (kWh)	MedAE (kWh)	MAAPE
Median	0.3745	**0.8964**	0.4120	0.6863
Average	0.3755	**0.9251**	0.4373	0.6867
Std	0.0440	0.1355	0.0780	0.0112
MLP ANN				
Error Metric	NRMSE	MAE (kWh)	MedAE (kWh)	MAAPE
Median	0.6884	1.0872	0.2644	0.5892
Average	3.3800	6.7540	0.2911	0.6103
Std	8.2716	20.0163	0.1166	0.1305

In summary, the proposed BN LF method resulted in better forecasting as indicated by the smaller NRMSE and the similar results for MedAE and MAAPE with the persistence method, which is in line with the errors presented in Fig. 5 and the forecasting illustration in Fig. 4. This reinforces the capability of the BN model to provide a coherent and generalised forecast for over the 900 residences used here. Besides these reported error measures, the computational times for our model learning and the execution of the VSTLF presented average values of 19 seconds, value observed on an i7@3.4 GHz desktop with

12 GB RAM and using bnlearn version 4.4 with R version 3.5.2.

A study that is directly comparable to ours is [18], where a BN was also used but for household demand prediction in a random divided data set without using the temporal sorting of consumption during prediction tests, i.e., this is not a load forecasting study. They reported a single NRMSE of 0.1182 when making an individual prediction for 25 households, and our model resulted in a median NRMSE of 0.0877 for over 900 households, a performance 25,8% better and using a data set 25 times bigger. It is important to remember that load time series at household level presents higher uncertainty and volatility compared to medium and high voltage power systems or aggregated load, making it an even more challenging forecast task, these factors become more relevant as it is a single BN model for hundreds of homes.

The challenge investigated in this study is related to forecasting individual household consumption based on a general model obtained by data-driven methods and historical data of consumption. We presented a single model that can handle the VSTLF of over 900 households without using their identification and using only four features that are related to households' socioeconomic electricity usage aspects together with the two past weeks of power consumption.

Other studies provided similar or even smaller forecasting errors than we presented here but at the cost of a large amount of data to model learning, i.e., use all available previous consumption data to learn a model and perform load forecasting, or models developed for individual building forecasting or also using information as household inhabitants routine or household identification. These peculiarities highlight aspects related to computational resources and scalability issues in the sense of forecasting a higher number of residences, and also privacy issues.

VI. Conclusions

This paper presented an approach to perform multiple households VSTLF without using load aggregation by a single generalist model together with multivariate data. A BN was used to model the relationship among the variables, which are past consumption measures, the households' electricity usage and socioeconomic aspects, and the temperature. The model VSTLF capability was evaluated for different days present in a real consumption data of 929 households and collected during a year and a half. It was also compared to other three forecasting methods: persistence, ANN and HMM.

As illustrated in the results, the proposed approach showed to be capable of generalising the power consumption patterns of multiple households. The proposed method presented better performance, as reflected in the 10 times smaller NRMSE together with the small differences for the other metrics when compared with the persistence, ANN and HMM. Furthermore, this satisfactory forecast for different instants needed only two weeks of past consumption data for model learning. This reinforces the approach capability of generalising the power consumption patterns of hundreds of households, allowing the forecasting of different residences with a single model.

In addition, the model also is computationally efficient, the experiments indicated that the model could be used in real time, as the learning and forecasting run-time are in order of seconds. This is related to the use of only two weeks of previous load consumption to learn the model. The space complexity is also low, as the model is composed of a directed acyclic graph together with the conditional tables related to the conditional dependencies requiring a small space to be stored. All this can allow the embedding of such an approach in smart meters or other "smart" components that can make use of the information provided in real-time by the forecasting.

It is important to highlight that the resulting model presented in this paper is specific to the set of households used in this study. However, since Bayesian networks are a data-driven approach, it should be applied to different data sets allowing an in-depth validation of our approach to multiple household VSTLF, which is one future research step, preferably using other publicly available household power consumption data. Future research also includes an optimisation of the variables necessary to perform the multiple household VSTLF, the exploration of the uncertainty around the estimation given by the current model, and also the investigation of BN use in other power systems forecasting problems, e.g., load forecasting in other time horizons and energy aggregation levels.

References

[1] S. Fallah, R. Deo, M. Shojafar, M. Conti, and S. Shamshirband, "Computational intelligence approaches for energy load forecasting in smart energy management grids: state of the art, future challenges, and research directions," *Energies*, vol. 11, no. 3, p. 596, 2018.

[2] T. Hong, J. Wilson, and J. Xie, "Long term probabilistic load forecasting and normalization with hourly information," *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 456–462, 2014.

[3] K. G. Boroojeni, M. H. Amini, S. Bahrami, S. Iyengar, A. I. Sarwat, and O. Karabasoglu, "A novel multi-time-scale modeling for electric power demand forecasting: From short-term to medium-term horizon," *Electric Power Systems Research*, vol. 142, pp. 58–73, 2017.

[4] L. Hernandez, C. Baladron, J. M. Aguiar, B. Carro, A. J. Sanchez-Esguevillas, J. Lloret, and J. Massana, "A survey on electric power demand forecasting: future trends in smart grids, microgrids and smart buildings," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 3, pp. 1460–1495, 2014.

[5] H. Shi, M. Xu, and R. Li, "Deep learning for household load forecasting—a novel pooling deep rnn," *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 5271–5280, 2018.

[6] T. Hong and S. Fan, "Probabilistic electric load forecasting: A tutorial review," *International Journal of Forecasting*, vol. 32, no. 3, pp. 914–938, 2016.

[7] Y. Wang, Q. Chen, T. Hong, and C. Kang, "Review of smart meter data analytics: Applications, methodologies, and challenges," *IEEE Transactions on Smart Grid*, 2018.

[8] F. Javed, N. Arshad, F. Wallin, I. Vassileva, and E. Dahlquist, "Forecasting for demand response in smart grids: An analysis on use of anthropologic and structural data and short term multiple loads forecasting," *Applied Energy*, vol. 96, pp. 150–160, 2012.

[9] M. Chaouch, "Clustering-based improvement of nonparametric functional time series forecasting: Application to intra-day household-level load curves," *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 411–419, 2014.

[10] A. Almalaq and J. J. Zhang, "Evolutionary deep learning-based energy consumption prediction for buildings," *IEEE Access*, vol. 7, pp. 1520–1531, 2019.

[11] D. Koller, N. Friedman, and F. Bach, *Probabilistic graphical models: principles and techniques*. MIT press, 2009.

[12] [dataset] Commission for Energy Regulation (CER), "Smart Metering Project - Electricity Customer Behaviour Trial, 2009-2010. 1st Edition. Irish Social Science Data Archive. SN: 0012-00," http://www.ucd.ie/issda/data/commissionforenergyregulationcer/, 2012.

[13] F. Kaytez, M. C. Taplamacioglu, E. Cam, and F. Hardalac, "Forecasting electricity consumption: A comparison of regression analysis, neural networks and least squares support vector machines," *International Journal of Electrical Power & Energy Systems*, vol. 67, pp. 431–438, 2015.

[14] R. R. Mohassel, A. Fung, F. Mohammadi, and K. Raahemifar, "A survey on advanced metering infrastructure," *International Journal of Electrical Power & Energy Systems*, vol. 63, pp. 473–484, 2014.

[15] F. Hafiz, B. Chen, C. Chen, A. R. De Queiroz, and I. Husain, "Utilizing demand response for distribution service restoration to achieve grid resiliency against natural disasters," *IET Generation, Transmission & Distribution*, vol. Accepted Manuscript, p. 12pp, 2019.

[16] W. Kong, Z. Y. Dong, D. J. Hill, F. Luo, and Y. Xu, "Short-term residential load forecasting based on resident behaviour learning," *IEEE Transactions on power systems*, vol. 33, no. 1, pp. 1087–1088, 2018.

[17] X. Fang, S. Misra, G. Xue, and D. Yang, "Smart grid—the new and improved power grid: A survey," *IEEE communications surveys & tutorials*, vol. 14, no. 4, pp. 944–980, 2012.

[18] N. Bassamzadeh and R. Ghanem, "Multiscale stochastic prediction of electricity demand in smart grids using bayesian networks," *Applied energy*, vol. 193, pp. 369–380, 2017.

[19] S. Kiartzis, A. Kehagias, A. Bakirtzis, and V. Petridis, "Short term load forecasting using a bayesian combination method," *International Journal of Electrical Power & Energy Systems*, vol. 19, no. 3, pp. 171–177, 1997.

[20] A. P. Douglas, A. M. Breipohl, F. N. Lee, and R. Adapa, "The impacts of temperature forecast uncertainty on bayesian load forecasting," *IEEE Transactions on Power Systems*, vol. 13, no. 4, pp. 1507–1513, 1998.

[21] P. Lauret, E. Fock, R. N. Randrianarivony, and J.-F. Manicom-Ramsamy, "Bayesian neural network approach to short time load forecasting," *Energy conversion and management*, vol. 49, no. 5, pp. 1156–1166, 2008.

[22] D.-x. Niu, H.-f. Shi, and D. D. Wu, "Short-term load forecasting using bayesian neural networks learned by hybrid monte carlo algorithm," *Applied Soft Computing*, vol. 12, no. 6, pp. 1822–1827, 2012.

[23] J. Pearl, *Causality*. Cambridge university press, 2009.

[24] R. E. Neapolitan *et al.*, *Learning bayesian networks*. Pearson Prentice Hall Upper Saddle River, NJ, 2004, vol. 38.

[25] M. Scutari, "An empirical-bayes score for discrete bayesian networks," in *Conference on Probabilistic Graphical Models*, 2016, pp. 438–448.

[26] E. Villanueva and C. D. Maciel, "Efficient methods for learning bayesian network super-structures," *Neurocomputing*, vol. 123, pp. 3–12, 2014.

[27] T. J. Gross, R. B. Araújo, F. A. C. Vale, M. Bessani, and C. D. Maciel, "Dependence between cognitive impairment and metabolic syndrome applied to a brazilian elderly dataset," *Artificial intelligence in medicine*, vol. 90, pp. 53–60, 2018.

[28] I. Tsamardinos, L. E. Brown, and C. F. Aliferis, "The max-min hill-climbing bayesian network structure learning algorithm," *Machine learning*, vol. 65, no. 1, pp. 31–78, 2006.

[29] F. Mallor, J. A. Moler, and H. Urmeneta, "Simulation of household electricity consumption by using functional data analysis," *Journal of Simulation*, pp. 1–12, 2018.

[30] [dataset] Irish Meteorological Service, "Irish meteorological history data," https://www.met.ie/climate/available-data/historical-data, 2018.

[31] G. I. Webb, "Contrary to popular belief incremental discretization can be sound, computationally efficient and extremely useful for streaming data," in *2014 IEEE International Conference on Data Mining*. IEEE, 2014, pp. 1031–1036.

[32] N. Ho, T. B. Pedersen, M. Vu, V. L. Ho, and C. A. Biscio, "Efficient bottom-up discovery of multi-scale time series correlations using mutual information," in *2019 IEEE 35th International Conference on Data Engineering (ICDE)*. IEEE, 2019, pp. 1734–1737.

[33] Y. Shi, C. Zhao, Z. Peng, H. Yang, and J. He, "Analysis of the lag effect of embankment dam seepage based on delayed mutual information," *Engineering Geology*, vol. 234, pp. 132–137, 2018.

[34] N. Ho, H. Vo, and M. Vu, "An adaptive information-theoretic approach for identifying temporal correlations in big data sets," in *2016 IEEE International Conference on Big Data (Big Data)*. IEEE, 2016, pp. 666–675.

[35] D. J. Albers and G. Hripcsak, "Using time-delayed mutual information to discover and interpret temporal correlation structure in complex populations," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 22, no. 1, p. 013111, 2012.

[36] T. M. Cover and J. A. Thomas, *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, 2006.

[37] R. Nagarajan, M. Scutari, and S. Lèbre, "Bayesian networks in r," *Springer*, vol. 122, pp. 125–127, 2013.

[38] M. Scutari and J.-B. Denis, *Bayesian networks: with examples in R*. Chapman and Hall/CRC, 2014.

[39] S. Kim and H. Kim, "A new metric of absolute percentage error for intermittent demand forecasts," *International Journal of Forecasting*, vol. 32, no. 3, pp. 669–679, 2016.

[40] S. Dutta, L. Y, A. Venkataraman, L. M. Costa, T. Jiang, R. Plana, P. Tordjman, F. H. Choo, C. F. Foo, and H. B. Puttgen, "Load and renewable energy forecasting for a microgrid using persistence technique," *Energy Procedia*, vol. 143, pp. 617 – 622, 2017.

[41] R. Rahmat, A. Rizki, A. Alharthi, and R. Budiarto, "Big data forecasting using evolving multi-layer perceptron," 2016.

[42] 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 E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[43] L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.

[44] D. Chambers, J. Baglivo, J. Ebel, and A. Kafka, "Earthquake forecasting using hidden markov models," *Pure and Applied Geophysics*, vol. 169, no. 4, pp. 625–639, 2012.

[45] A. Lahouar and J. B. H. Slama, "Day-ahead load forecast using random forest and expert input selection," *Energy Conversion and Management*, vol. 103, pp. 1040–1051, 2015.

[46] X. Qiu, Y. Ren, P. N. Suganthan, and G. A. Amaratunga, "Empirical mode decomposition based ensemble deep learning for load demand time series forecasting," *Applied Soft Computing*, vol. 54, pp. 246–255, 2017.

[47] I. Koprinska, M. Rana, and V. G. Agelidis, "Correlation and instance based feature selection for electricity load forecasting," *Knowledge-Based Systems*, vol. 82, pp. 29–40, 2015.

[48] The Irish Times, "December was coldest month in 28 years," https://www.irishtimes.com/news/december-was-coldest-month-in-28-years-1.851279?via=rel, 2010.

[49] Independent, "Sub-zero temperatures make 2010 a record-breaking year," https://www.independent.ie/irish-news/sub-zero-temperatures-make-2010-a-record-breaking-year-26609480.html, 2010.