Assessing Error Bars In Distribution Load Curve Estimation

J. N. Fidalgo jfidalgo@bart.inescn.pt

M.A. Matos mmatos@inescn.pt

M. T. Ponce de Leão mleao@inescn.pt

Abstract: Electrical distribution utilities have been dealing with the problem of estimation of distribution network load diagrams, either for operation studies or in forecasting models for planning purposes. Load curve assessment is essential for an efficient management of electric distribution systems. However, the only information available for most of the loads (namely LV loads) is related to monthly energy consumptions. The general procedure uses measurements in consumers to construct inference engines that predict load curves using commercial information.

This paper presents a new approach for this problem, based on Kohonen maps and Artificial Neural Networks (ANN) to estimate load diagramsfor the portuguese distribution utilities. A method for estimating error bars is also proposed in order to provide a high order information about the performance of load curve estimation process. Performance attained is discussed as well as the method to achieve confidence intervals of the main predicted diagrams.

Keywords: load estimation, fuzzy clustering, Kohonen maps, ANN, confidence intervals

1 Introduction

The two last decades had testified a continuous growth on consumptions' estimation studies [1-5]. Several distribution utilities performed these studies, modelling consumers' behaviour for planning purposes, and using inference processes usually based on linear multiregression.

Aiming at solving network planning needs, EDP (Portuguese Power Company) in cooperation with INESC, developed a Load Curve Management Module that estimates hourly load diagrams for the distribution network. The project included measurement campaigns, model development and testing. EDP has carried out the measurement campaigns, collecting consumptions' evolution data, in order to implicitly characterise consumers' behaviour. A large spectrum of possible load curves is crucial to represent the whole universe of consumers.

The adopted approach can be divided in three phases: a) definition of consumers' classes (clustering) from general and commercial information; b) inference of hourly consumptions of a given consumer, based on its classification and monthly energy consumption; c) estimation of error bars. On phase a), Kohonen maps were used in the clustering process. Step b) require the training ANN's for multiregression purposes, in this case, to obtain load curves from commercial data. ANN's used for step c) are trained to learn the error presented by ANN's obtained on b).

The following sections describe the objectives of all these phases. Finally, some results for LV cases are presented and commented.

2 Characterization and objectives

2.1 Clustering

Consumers' classes were defined solely from consumers' load curves. This approach has the advantage of producing clusters based on the actual evolution of loads. Each clustering training pattern contains 24 elements - the registered power consumptions at each hour of a given day. This training set was presented to a Kohonen clustering tool [6], in order to obtain different load evolution classes. Results were compared with other classification tools namely Bezdek's Fuzzy Isodata algorithm [7].

2.2 Load Curve Estimation

When dealing with MV networks, the main issue is to get load figures consumptions, in any point of the network, to be used later for planning purposes. The available information from EDP data base consists mainly on commercial data and energy consumptions.

One intends to estimate load diagrams essentially for MV/LV public stations and MV individual clients. It was also decided to aggregate LV consumers, dependent from public stations, in order to evaluate their accumulated load diagram. This will avoid the need for the characterisation of each LV individual consumer, reducing the size of the data base needed for future studies. Furthermore, there are no imperative knowledge requirements for a single LV client.

The data obtained from the measurements campaigns was divided following the season (Summer or Winter), the weekday (workday or weekend) and the type of consumer (LV or MV). Other available data of LV consumers are the monthly energy consumption, the activity code and hired power. All those consumers are fed by MV/LV public substations. For MV consumers, the accessible activity code, peak power, hired power and energy measures for different tariffs are known.These curves and values must be available in several points of the network, for instance in a MV/LV public station or in a feeder. Additional parameters are evaluated for each load curve (peak power use, load factor, loss factor, etc.), which help the characterisation of single or aggregated consumer behaviour.

3 MODELS

Consumers' modelling needed two different types of analysis: MV clients and MV/LV public stations. In fact, a lot of public stations have no load measurement at all, and it is not possible to infer directly its hourly consumption. Moreover, as a result of the different behaviour according to time and season of the year, the analysis was divided into Winter/Summer and week/weekend day cases. The establishment of the partial models for all the mentioned cases was followed by the development of a integration procedure to cope adequately with intermediate situations. Due to space limitations only the LV case will be presented.

3.1 Class definition - clustering

Some electrical distribution utilities define an *a priori* classification of consumers based on its commercial characteristics. Here, one of the fundamental steps of the adopted approach was the identification of natural classes from registered diagrams. For that purpose, two different methods were tried: the Fuzzy Isodata algorithm and self-organised neural networks (Kohonen maps).

After obtaining the cluster's groups one's must to induce the relation between classifications and commercial data (tariff class, hired power and monthly consumption), in order to generalise the classification of consumers for which only commercial data is known, (what constitutes the real future operational conditions).

Some experiences were carried out for determining a good combination of clustering (number of classes based on load evolution) and inference of classification rules (based on commercial data). This inference process was based on the observation of the distribution of classes's members on the 3D space of commercial data (tariff class, hired power and monthly energy consumption). The best classification performance was obtained with four classes (figure1) and the following classification rules:

- **A** Domestic consumers (T_c=0), low hired power (P_c≤6.6kW) and low energy consumption (E<600kWh);
- **B** Industrial consumers (T_c=4);
- **C** Other consumers (except night consumers);
- **D** Night consumers.

This was validated through a detailed analysis to the list of diagrams belonging to each class. One's must stress that Fuzzy Isodata algorithm produced cluster prototypes very similar to Kohonen ones.

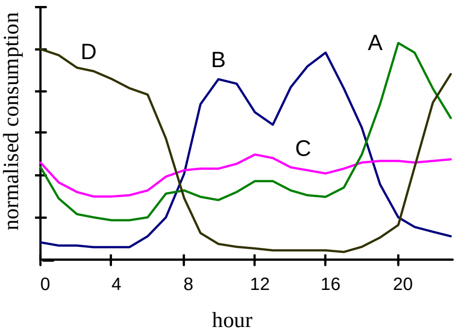


Figure 1 - Kohonen cluster prototypes

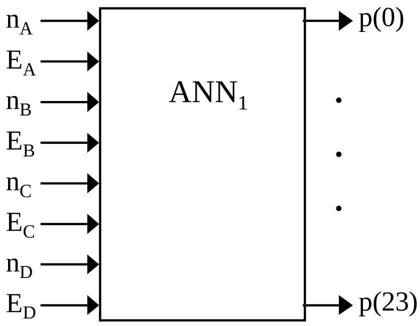


Figure 2 - Inputs/Outputs from ANN

3.2 Diagram inference

For the **Public stations** case, the available data for each one of these stations only comprises:

- Number of LV consumers in each operational class;
- Total consumption of energy (monthly) of each class.

The daily diagram estimation in an hourly base (p_0, p_1, ..., p_{23}) is made using a back-propagation neural network, as the one, called ANN_1, represented in figure2.

n_i and E_i are respectively the number of consumers and the total energy (monthly) of class i.

To train this ANN, 2000 patterns were generated from the data file derived from consumption measurement campaign. Each pattern was generated to include from 80 to 160 LV consumers, randomly selected from basic sample. For each pattern, the 8 ANN values derive from the classification of prototype elements, followed by energy counting and sum for each class. The values from the outputs have equivalence in the 24 time intervals from the aggregated diagram of the consumers belonging to each pattern. From the 2000 generated patterns, 1500 were taken out to train the ANN and the remaining were used for test. Figure 3 presents pattern examples of the test set, comparing the real diagram (cons.) and ANN outputs (NN). The exemplar ANN has an input/output structure presented in figure2 and was trained with data of LV consumer, summer and workdays.

Figure 4 similar results but for winter workdays. Global results show that the ANN is capable of estimating the test set diagrams presenting a rms error around 10%. This may be considered a good result, if we take into account the arbitrariness inherent to loads evolution.

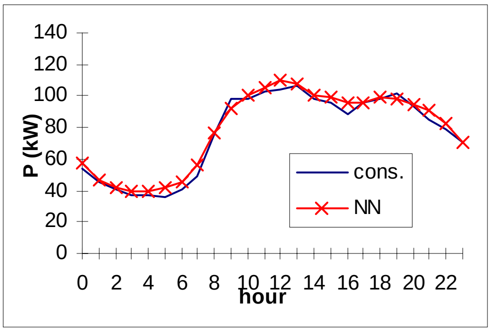


Figure 3 - Inference test (LV Summer workday)

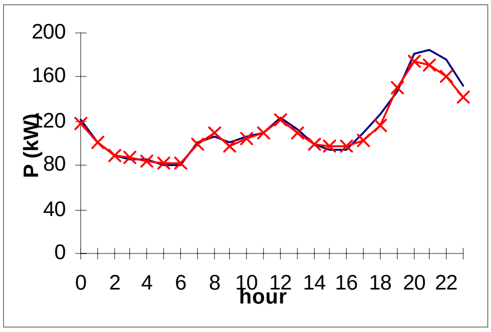


Figure 4 - Inference test (LV Winter workday)

4 Confidence intervals

Despite the good quality of approximation achieved (figures 4 and 5), it is always desirable to access some measure of the accuracy of each estimated load curve. Of course ANN's average errors are always available but they are computed for all the pattern set; the specificity of consumers characteristics is not taken into account, i. e., average effects will hide the differences between the very good performance of some estimates and the poor performance of other estimates. Electrical distribution planners would prefer to obtain a given bandwidth around ANN_1 estimated load diagram, in such way that the probability that a real load curve be inside that band is, let's say, 0.9.

Figure 5 presents aggregated load diagrams in two given public stations (each graphic) on different week days (each grey curve inside a graphic). The analysis of a large amount of figures like this one has shown that there is a structure on the errors spreading. If there is not, we can only evaluate average errors. We can observe that consumptions' dispersion is not homogeneous, that is, the same consumer or group of consumers does not present the same uncertainty around a medium load curve for all the hours of the days. For instance, the dispersion before 6 a.m. is usually smaller than at (*e.g.*)11:00. There has been some interesting work in the area of confidence interval prediction for ANN's [8-10]. In most of those studies, authors assume Gaussian or t-student distributions and estimate output variance as a function of inputs variance and of input/output transferring function, using Bayes rule. Some of these assumptions are discussed in [11].

On the present work, there was no need to make any assumptions about distributions of ANN_1 inputs, outputs, weights or errors. We propose to train an auxiliary ANN (called ANN_d and shown in figure 6) to learn load curves's dispersion (error bounds distribution) around ANN_1 output. These error bars will depend not only on the type and number of consumers aggregated in a given public station, but also on the hour of the day. Inputs of ANN_d are the same of ANN_1. Its outputs are the absolute values of the differences between ANN_1 outputs and load consumption curves. This way, ANN_d produces an error dispersion measure of diagrams estimated by ANN_1.

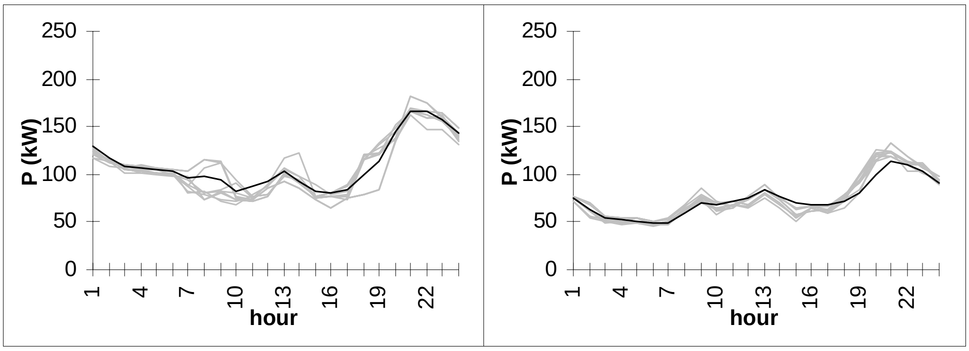
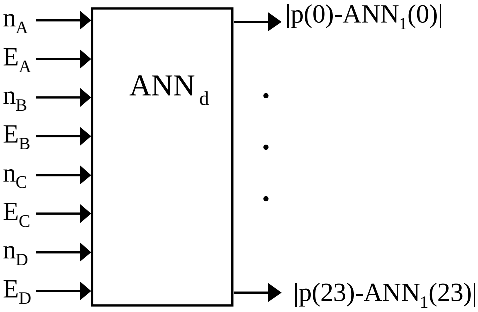


Figure 5 - Two examples of aggregated consumption (grey) and ANN estimation (black); within each graphic there are several curves for a given public station for different workdays



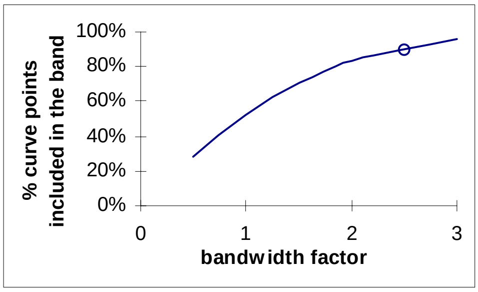


Figure 6 - ANN for estimation of error bars

Figure 7 - Inclusion factor versus bandwidth

It must be stressed that there are two kinds of errors: a) errors arising from ANN implementation limitations and b) errors (called dispersion errors) related with the nature of predicted values. We shouldn´t expect that ANN_d learn the approximation errors that ANN_1 couldn't learn, but we hope that learn something about the way total error distributes itself over ANN_1 outputs and as a function of its inputs.

After training, the outputs of ANN_d were computed for each pattern of the training set and a error bands were defined around each predicted diagram. Then, for each pattern, we evaluate the percentage of load curve points inside its band. We proceed with this study by repeating the calculus with ½*band, with 2*band, and so on. As a result of this study, Figure 7 was sketched. This figure characterises the relation between what we have called bandwidth factor and the percentage of hourly consumptions within the bands. One's can observe the 90% confidence intervals corresponds roughly to 2.5 times the band error (output of ANN_d). Figure 8 is a repetition of Figure 6 but also represents the 90% confidence bars around ANN_1 output. This tool provide power system plannes a measure of the confidence interval of the estimated load curve.

5 Conclusions

It was generated an inference mechanism to estimate consumption from available commercial data base and monthly energy consumptions for MV/LV public stations and MV consumers.

Results obtained show the adequacy of the adopted approach.

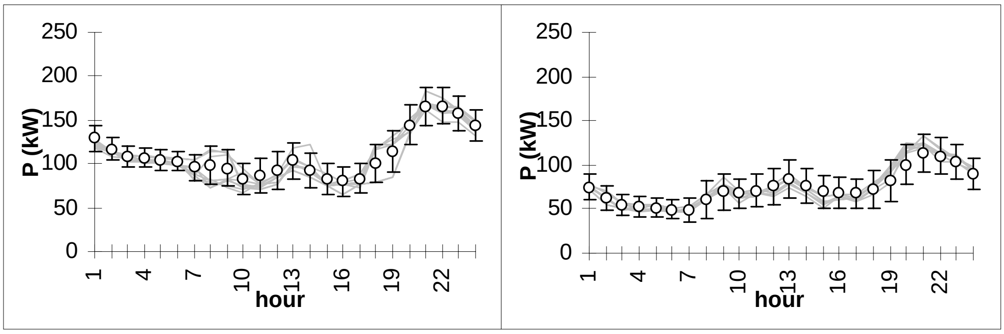


Figure 8 - Similar to figure 6 but showing the 90% confidence intervals

The estimation of confidence intervals was performed by a simple but efficient strategy: using a second ANN to access the error of the first one. This approach has the capital advantage of including all kind of errors inherent to the load estimator ANN. Moreover, we don't have to make any kind of assumptions about training data or weights distribution, or to force any other premises in order to use Bayes theorem.

The work made so far leads to the conclusion that adopted approach is able to bring very interesting results and can facilitate adequate results for planning purposes. Presented results contribute to confirm that adopted tools are most suitable for the proposed objectives.

Referencees

[1] Comellini, E. et al., "Correlations entre Puissance et Energie Consommée par les Charges des Reseaux Publiques de Distribution", *Proceedings CIRED*, Liège, 1979

[2] Canal, M., Fieux, H., Landry, M., "Le sistème informatique de gestion des ouvrages BT. Calcul des charges appelées par les réseaux basse tension", *RGE*, Nº2, 1984

[3] Pensoz, H., et al., "La planification des réseaux éléctriques", *Planification des Réseaux Moyenne Tension Eyrolles*, 1984

[4] Juuti, P., Lakervi, E., Partanen, J., "The Use of Customer Load Profiles in Distribution Network Design and Operation Planning", *Proceedings ISEDEM*, Singapore, 1988

[5] Oliveira, L., *et. al*., "Load Curve Management", *CIRED* Argentina, 1996

[6] Kohonen, T., *Self Organization and Associative Memory*, Springer-Verlag, Berlin, 1984

[7] Bezdek, J.C., *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New York, 1981.

[8] Nix, David A. and Weigend, Andreas S., "Learning Local Error Bars for Nonlinear Regression", In *Advances in Neural Information Processing Systems*, G. Tesauro, D.S. Touretzky, and T.K. Lean (Eds.), MIT Pess, 1995

[9] Chryssolouris, George, "Confidence Interval Prediction for Neural Networks", in *Trans. on Neural Networks*, Vol. nº. 1, Jan 1996

[10] Tom Heskes, "Practical Confidence and Prediction Intervals", In Advances in Neural Information Processing Systems, M. Mozer, M. Jordan, T. Petsche (Eds.), MIT Press, 1997

[11] Bellido, I. and Fiesler, E., "Do Backpropagation Neural Networks have Normal Weight Distributions ?", *ICANN'93*, 1993