# **Forecasting Daily Electricity Load Curves**

RUI COELHO <sup>(1)</sup> and ANTONIO J. RODRIGUES <sup>(2)</sup> <sup>(1)</sup> REN – Rede Electrica Nacional, Sacavem, PORTUGAL

<sup>(2)</sup> DEIO-FCUL and Centro de Investigação Operacional, Universidade de Lisboa, 1749-016 Lisboa, PORTUGAL

*Abstract:* Short term electricity load forecasting is a well-known problem, and many neural computing approaches for solving it have been proposed in recent years. In this paper, we argue in favour of its decomposition into two subproblems, and propose a solution for one of them: the prediction, *en bloc*, of the daily load profile, or configuration, for the different hours of a particular future date. From this point of view, we propose a methodology where the *shape* of the load curve is inferred from two variables: one of them is nominal and serves to characterize the type of day, calendarwise; the other one consists of a pattern of temperature forecasts for that day. The approach includes fuzzy clustering of past temperature patterns, as well as fuzzy clustering of past load curves, and inference is based on the computation of empirical correlations among the three variables: the day class, the temperature pattern, and the load pattern.

Key-Words: - Load Forecasting, Neural Networks, Fuzzy Clustering, Pattern Recognition, Classification.

# **1** Introduction

One of the most important missions of electricity transport system operators is to produce and provide forecasts of electricity consumption. On the quality of the forecast depends the efficient balance between production and demand, which has a direct impact on the safety of the electrical system.

In the last few years, many countries, including Portugal, have privatised and deregulated their power industries, turning electricity into a kind of commodity to be bought and sold at a market price. With the rise of a competitive market, the load forecast problem has become increasingly popular and important, as accurate short-term demand forecasting will also be required by those who will trade electricity on financial markets.

Short term load forecasting (STLF) usually refers to time horizons of just a few hours up to a few days ahead, from past data collected on a hourly or subhourly basis. The aim is to predict future electricity daily load patterns based on the recognition of similar repeating patterns in the historical data. This goal must also take into consideration important exogenous factors, including calendar effects, and weather-related variables, most importantly the temperature.

The research on the STLF problem has developed considerably in the last 10-15 years, with a significant increase in the number of proposed solutions based on or related to Neural Computing. This can be noticed in the contents of some journals, such as the *IEEE Transactions on Power Systems*, in many conferences, in new specialized commercial forecasting software, and in international forecasting competitions dedicated to the problem. The most common approaches include supervised ones namely, Feedforward or Recurrent Multilayer Perceptrons and Radial Basis Function Networks —, as well as unsupervised ones — namely, Kohonen's self-organizing maps. Recent surveys of the literature include [2], [3], [5] and [7].

Neural computing approaches may be better suited for the STLF problem than simpler, more conventional, linear models, if they are used to exploit the intrinsic pattern recognition and nonlinear correlation aspects of that problem. However, even if no exogenous variables are considered, modelling hourly load time series through supervised neural networks as nonlinear autoregressive processes can be very cumbersome or ineffective. On the one hand, single-output networks, designed for iterative one-hour-ahead prediction typically lead to poor forecasts at the end of a 24-hour or longer forecasting horizon. On the other hand, a network with 24 outputs, representing the hourly forecasts for a whole day, would be a very complex model to be identified and estimated, with a large number of nonlinear parameters [4].

Therefore, many authors seek to reduce the complexity of the problem by using Cluster Analysis, Principal Component Analysis, or other data reduction techniques (see, e.g., [6]).

In this paper, we suggest decomposing the STLF problem into two less complex ones:

- predicting the shape of daily load curves, irrespectively of location and scale features;
- predicting daily location and scale features.

Forecasting the whole daily load curve is much more useful and informative that just forecasting the daily peak demand. And, for planning purposes, it is often important to have equally accurate forecasts for the different hours of a given future date.

In the following Section, we discuss the main features of the load data from the Portuguese transport system operator, which are very similar to those of other countries. In Section 3, we present our approach to solve the first of the above two subproblems, and in Section 4 some sample results with the Portuguese data are reported.

### 2 A Case Study

We consider data from REN-Rede Eléctrica Nacional, the Portuguese transport system operator. The data refers to the instantaneous values, in MW, of the half-hourly total load on the Portuguese grid for years 2001 to 2004.

As the load process is strongly related with the human activity cycles, we can identify, apart from a linear trend, three main seasonal effects in the data: an annual cycle — with a peak in December/January and a slack period in the middle of August (Fig. 1) —, a weekly cycle — showing a generally stable demand during the weekdays and a drop on weekends (Fig.2), and a daily cycle (Fig. 3). In this paper, we focus our attention in the latter.

This regular behaviour is however disrupted by the occurrence of holidays, and days of other special events. Fig. 4 shows the monthly load pattern of December 2004. The load shape becomes anomalous due to three holidays: December 1st, Dec. 8th and Christmas Day.

The load data can be splitted into daily segments, or *load curves*, which exhibit somewhat similar but sufficiently different profiles. Those differences are very important, as they reveal when maximum instantaneous demand is attained, for how long, etc. This segmentation is of great value because it enables a best suited form of describing the process in the sub-daily time scale.

As can be seen in Fig. 3, there is an obvious difference between the shapes of the load curves on a typical working day, on a Saturday, or on a Sunday. Furthermore, the load shape depends not only on the day of the week but is also related with the time of the year.



Fig.1 - Weekly averaged load in 2003 and 2004



Fig.2 - A typical weekly load pattern (day1=16 Oct 2004, Monday)



Fig.3 – Load profiles from a working day (solid line), a Saturday (dashed line) and a Sunday (dotted line) of October 2004



Fig.4 - Monthly load pattern of December 2004

For instance, the typical working days in the Winter exhibit a higher evening peak due to the superimposition of the labouring hours with the night period (Fig. 5).

Temperature is responsible for the main differences in load shapes, given comparable calendar days. Fig. 6 depicts the load profiles of two consecutive and similar Thursdays in 2004. On those days temperatures were in the range 5.1-10.5°C and 10.3-18°C respectively. It should be noted, however, that there is a nonlinear relationship between load and temperature (Fig. 7), which is variable with the time of the year, and that these variations include changes in magnitude and sign.



Fig.5 Load profiles from a Winter working day (solid line) and a Summer working day (dotted line)







Fig.7 Scatter plot of temperature-load relationship for working days in 2004.

# **3** Proposed Approach

We assume the availability of two univariate time series, recording actual loads and temperature forecasts for the past N days. For each day, loads are recorded at regular intervals — in our case, 00h00, 00h30, 01h00, ..., 23h30 —, and temperature forecasts are available for different hours of the day — e.g., 00h00, 06h00, 12h00 and 18h00.

Furthermore, one defines *a priori* a classification for the days into *ND* possible classes, according to calendar criteria, taking into account the time of the year, the day of the week and the different holidays.

#### 3.1 Divide to Conquer

First, we define the series of *N* daily load curves and temperature "curves":

 $\{\mathbf{y}_t\}_{t=1,..,N}$ : normalized patterns of 48 load values;

 $\{\mathbf{x}_t\}_{t=1,\dots,N}$ : patterns of 4 temperature forecasts.

Since, in the present paper, we focus mainly on the problem of predicting the *shape* of the daily load curve, ignoring location and scale features, the patterns  $\{\mathbf{y}_t\}$  are normalized to have zero mean and unit variance.

The original means and standard deviations,  $\{m_i\}$ and  $\{s_i\}$  are two univariate time series, to be modelled and forecasted separately. These are clearly nonstationary, and one is advised to prefilter each of them with a model for the trend and annual seasonal effects before resorting to a supervised neural network to account for the weekly seasonal effects, the holidays and other anomalous days, and the nonlinear effect of the temperature. A relatively small network could then be used for that purpose:

- Inputs:
- information related to a few past days (*m* and s; type of day; temperature);
- information related to a future date (type of day; temperature forecasts);

Outputs:

-m and *s* for that day.

Supposing one wished to compute *p*-steps-ahead forecasts of the load for a future date  $(p \ge 1)$ , given the 4 temperature estimates for that day,  $\mathbf{\hat{x}}_{N+p|N}$ , the predicted load curve may then be found simply by scaling  $\mathbf{\hat{y}}_{N+p|N}$  with  $\hat{s}_{N+p|N}$  and by shifting with  $\hat{m}_{N+p|N}$ . To render the final forecasts more robust and accurate, it is advised to model and forecast as well the univariate time series of loads at, say, 06h00 and 18h00, and make appropriate adjustments for consistency.

### 3.2 Clustering

Since the aim is to try to infer an estimate of the load curve for a future day, knowing its class and the temperature forecasts, one needs to determine how do "type of day" and temperature jointly influence the load. Load profiles and temperature profiles cannot be classified but they can be clustered.

The classification into types of day is made *a priori*, and is therefore indisputable, but clustering is considered intrinsecally ambiguous and uncertain. However, we can quantify and build upon that uncertainty by resorting to a fuzzy clustering method [1] and by computing, for each load or temperature pattern, its *degree of membership* to the previously identified clusters. These values, if normalized, can be losely regarded as pseudo-probabilities of a pattern belonging to each of the different clusters. Another approach to compute cluster membership probabilities would be to base the clustering on a mixture model.

Fuzzy clustering is then applied to the  $\{\mathbf{y}_t\}$  patterns, and the respective *NY* centres and degrees of membership,  $\{p_{tk}\}$ , determined:

$$\sum_{k=1,...,NY} p_{k} = 1, \quad (t = 1,...,N)$$
(1)

Similarly, fuzzy clustering is applied to the  $\{\mathbf{x}_t\}$  patterns, and the respective centres and degrees of membership,  $\{q_{ii}\}$ , determined:

$$\sum_{j=1,...,NX} q_j = 1, \quad (t = 1,...,N)$$
(2)

In this case, the  $\{\mathbf{x}_t\}$  patterns are not normalized — on the one hand, it would not be reasonable to compute standard deviations; on the other hand, we should take into consideration the influence of the magnitude and amplitude of the temperature values on the load profiles.

### 3.3 Empirical Inference

We then define a 3-dimensional matrix,  $\mathbf{F} = \begin{bmatrix} f_{ijk} \end{bmatrix}$ (*ND*×*NX*×*NY*), of estimated pseudo-correlations among the three variables: type of day, temperature and load. If crisp clustering were used, cell  $f_{ijk}$ would indicate a frequency, namely the number of past days belonging to class *i*, whose load pattern was included in cluster *j*, and whose temperature pattern was included in cluster *k*. With fuzzy clustering, we have instead:

$$f_{ijk} = \sum_{t \in D(i)} p_{ik} q_{ij}$$
(3)

where D(i) denotes the subset of days of class *i*.

Given the pattern  $\hat{\mathbf{x}}_{N+p|N}$ , and if  $N+p \in D(i)$ , one might consider the following simplistic procedure: find the closest temperature cluster, *j*, then determine

$$\ell: f_{ij\ell} = \max_{k} \left\{ f_{ijk} \right\} \tag{4}$$

and use the center of the load cluster  $\ell$ ,  $\mathbf{c}_{\ell}$ , as the predicted load profile. But that would be too myopic, and prone to produce significant prediction errors.

Instead, we compute the degrees of membership of  $\hat{\mathbf{x}}_{N+p|N}$  to the temperature clusters,  $\{q_{N+p,j}\}$ , and define

$$\hat{\mathbf{y}}_{N+p|N} = \sum_{j=1,\dots,NX} q_{N+p,j} \,\hat{\mathbf{z}}_{N+p,j} \tag{5}$$

where each  $\hat{\mathbf{z}}_{N+p,j}$  is an estimate of the load profile of day N+p conditional to the temperature profile of that day 'belonging' exactly to cluster *j* (and to no other).

Each  $\hat{\mathbf{z}}_{N+p,j}$  is estimated by linearly combining all the centres of the load clusters, according to the values  $f_{ijk}$ , where *i* is the class of day N+p:

$$\hat{\mathbf{z}}_{N+p,j} = \sum_{k=1,\dots,NY} w_{ijk} \, \mathbf{c}_k \tag{6}$$

where

$$w_{ijk} = f_{ijk} / \sum_{\ell=1,\dots,NY} f_{ij\ell} .$$
<sup>(7)</sup>

# **4** Experimental Results

The Portuguese data, introduced earlier, was splitted into two subsets. A subset, corresponding to years 2001-2003, was used for clustering and for the computation of matrix **F**. The 2004 data was used as a validation set, namely to support the choice of the number of clusters, by minimizing the average (over the year) root mean squared *fitting* error in predicting the next daily curve of 48 points. We have set NY=24 load clusters and NX=10temperature clusters, and noticed that close, but not much lower values produce similar performance.

Moreover, we have considered a total of *ND*=41 classes, including, for instance: Spring Sunday; Summer Monday; Autumn weekday (except Monday); Easter; etc.

Instead of past records of temperature forecasts, we resorted only to past actual values of the temperature, and have tested the approach assuming the four temperature records of one day (at 00h00, 06h00, 12h00 and 18h00) were used as the forecasts for the following *days*. Naturally, one should achieve better performance by using proper forecasts, especially for multiple-day horizons.

Since this paper is focused on the prediction of the shape of normalized load curves, we report only results in this respect, and only for 1-day-ahead forecasts. For the sake of illustration, in Fig. 8 we compare the predicted with the actual (normalized) curves on a sample day, for which the root mean squared fitting error (RMSE) was 0.165. The RMSE values for the 366 days of year 2004 are shown in Fig. 9, and the average was as low as 0.175.

### **5** Conclusions

The complete neural-based forecasting of future loads, in the original measurement scale (MW), is greatly facilitated by the solution introduced in this paper. We proposed a practical and efficient procedure to forecast, *en bloc*, the shapes of daily load patterns, one or more days ahead, that takes into account calendar and temperature effects. Test results were very promising, and there is room for further improvement. The methodology can be easily extended to more dimensions, to accomodate other possible influential variables, and can also be easily adapted to other demand forecasting problems.

A crucial aspect to the success of the proposed approach is the sensible *a priori* classification of days into (many) meaningful types related to the calendar, which can even be further refined. Other criteria, related to special scheduled events, might be considered as well. Improved results are also expected if more than 3 years of past data are considered, thus enabling a better characterization of the classes with fewer instances.

Another key aspect of the methodology was the use of a soft clustering approach to the identification of clusters for load patterns and temperature patterns. This has made possible the computation of degrees of membership, and the combination of multiple centres in the estimation process. The resulting estimates are then fairly robust, as they are actually based on *all* past data available.



Fig.8 Estimated (dotted line) vs actual load (solid line) for 25.01.2004





References:

- Balasko, B., Abonyi, J., Feil, B., *Fuzzy Clustering* and Data Analysis Toolbox For Use with Matlab, Department of Process Engineering, University of Veszprem, Hungary, 2005.
- [2] Bansal, R.C., Pandey, J.C., Load Forecasting Using Artificial Intelligence Techniques: A Literature Survey, *International J. of Computer Applications in Technology*, Vol.22, No.2/3, 2005, pp. 109-119.
- [3] Crone, S.F., *Neural Forecasting Bibliography*, Centre for Forecasting, Lancaster University, 2005, http://www.neural-forecasting.com.
- [4] Fay, D., Ringwood, J., Condon, M., Kelly, M., 24-Hour Electrical Load Data – A Sequential or Partitioned Time Series?, *Neurocomputing*, Vol.55, 2003, pp. 469-498.
- [5] Hippert, H.S., Pedreira, C.E., Souza, R.C., Neural Networks for Short-Term Load Forecasting: A Review and Evaluation, *IEEE Transactions on Power Systems*, Vol.16, No.1, 2001, pp. 44-55.
- [6] Lendasse, A., Lee, J., Wertz, V., Verleysen, M., Forecasting Electricity Consumption Using Nonlinear Projection and Self-Organizing Maps, *Neurocomputing*, Vol.48, 2002, pp. 299-311.
- [7] Metaxiotis, K., et al, Artificial Intelligence in Short Term Electric Load Forecasting: A State-of-art Survey for the researcher, *Energy Conversion and Management*, Vol.44, 2003, pp. 1525-1534.