Instituto de Engenharia de Sistemas e Computadores de Coimbra

Institute for Systems and Computers Engineering

INESC – Coimbra

João C. Sousa, Luís P. Neves, Humberto M. Jorge

Integrating information available from load profiling in electric load forecasting

No. 17 2009

ISSN: 1645-2631

Instituto de Engenharia de Sistemas e Computadores de Coimbra

INESC – Coimbra

Rua Antero de Quental, 199; 3000 - 033 Coimbra; Portugal

www.inescc.pt

Integrating information obtained from Load Profiling in Electric Load Forecasting

João M. C. Sousa (1) , Luís M. P. Neves (1) , Humberto M. M. Jorge (2)

(1) Dept. of Electrical Engineering, Campus 2, Polytechnic Institute of Leiria,
2411-901 Leiria, Portugal

(2) Dept. of Electrical Engineering and Computers, Pólo II, University of Coimbra,
3030-290 Coimbra, Portugal

(1), (2) Institute for Systems Engineering and Computers – INESC Coimbra,
Rua Antero de Quental, 199, 3000-033 Coimbra, Portugal

December 2009

***Abstract*:** Recent researches in load forecasting are quite often based on the use of neural networks in order to predict one record (single output) or more records (multiple output) of a specific variable (normally related to maximum demand, hourly-based active electric power or hourly consumption) using past values of the same variable and other exogenous variables, proved to influence the value being predicted. Besides the consensual influence of variables such as meteorological variables, seasonal effects or macroeconomic indexes, it is expected that the use of typical load profiles and detailed information of individual consumers could favour the forecasting process. Therefore, this work aims to explore different input patterns in neural networks architectures, incorporating information derived from load profiles of different classes. The methodology can be extended to different temporal horizons being predicted and the eventual threat of overparametrization is attenuated by the use of neural networks, since the complexity of the model does not necessarily depends on the number of its weights and biases.

Keywords: Load Forecasting, Load Profiling and Neural Networks

1. Introduction

The deregulation of electric power markets has effectively brought new and varied types of challenges, being actually well proven that special attention has being paid to issues such as load forecast and electricity consumers classification.

Load forecasting has always been understood as a key issue in the electricity sector, since it helps to provide a physical balance between the supply and the demand, being essential in order to support an analysis of an eventual strengthening or expansion of the existing infrastructure and implementation of a maintenance scheduling (in a long-term management to guarantee a reliable system operation). It is also an important tool to provide an optimized network configuration, being valuable to unit commitment or increasingly more helpful to plan the integration of dispersed generation (in a short-term perspective), knowing that electric energy must be always available at any period of time and at a minimum price.

Assisting to the worldwide trend of the sector liberalization and the subsequent unbundling of the value chain activities, load forecasting achieves even more importance, not only for system operators but also for any other participants, to schedule adequately energy transactions. The minimization of forecasting errors is crucial as they would result in increased operational costs and loss of revenue. Several types of methods have been applied to forecast future demand, such as time series models (Campbell (2006), Chen (2001), Hippert (2001), Willis 2004)) (using linear, polynomial or exponential regression), based on previous records of the variable being predicted and/or based on exogenous factors that can directly or indirectly affect the load behaviour, in which the variable can be estimated as a function of meteorological conditions, economic activity, macroeconomic indexes (consumer price index or average salary earning), day type or electricity price. Recent load forecasting research adopts artificial intelligence and the neural networks models are undoubtedly the most popular strategy in this context (Campbell (2006), Chen (2001), Hippert (2001)).

On the other hand, the consumer's classification is actually mandatory for many electricity markets, being crucial for enabling the market participation of all consumers, even those who have not installed hourly-metering systems. The use of load profiles for each consumer's group characterizes different kinds of electricity usage and makes the settlement between distributors and suppliers possible, as the total energy record available in the traditional meters can be distributed through different interval periods. The information given by load profiles may be useful to retail companies as it helps to establish contracts with different kinds of consumers, with specific characteristics, practicing adjustable prices for classes and offering added-value services adjustable to each customer group. In the distributor's perspective, this detailed information may also have relevance in load management and in electric network access tariffs definition.

This work intends to evaluate the eventual contribution of consumer classification in load forecasting thematic, adopting neural networks to capture relationships between the variable(s) being predicted and some specific information (inputs in the forecasting model) known to have strong correlation with the previously unknown variable(s). The main objective is to integrate load profiles information in the input variable space, evaluate the methodology and comment the obtained results.

This report is organized according to the following description: the next section will describe in more detail the load forecasting methods, focusing on neural networks. Section 3 is dedicated to load profiling, advantages and motivation to this strategy and also to synthesize different ways used to segment electricity consumers based on the energy use. In Section 4 the case study will be presented with detailed description of different stages associated with the proposed methodology. Section 5 presents some obtained results and also includes comments and comparisons between different forecasting methods used. The last section points out the principal conclusions and topics for further work.

2. Load Forecasting

The use of artificial neural networks (ANN) in load forecasting is actually pointed out as an alternative or, in some cases, a complement to the traditional time series models. Falling into the group of artificial intelligence techniques, the ANNs have been explored since the middle 80's, with tendency to gain more adepts and proving its accuracy to some sceptic researchers (Hippert (2001)). Some of the principal advantages often commented are its ability to model nonlinear and complex relationships better than traditional linear models, providing an easy way to deal with multivariate models and exploring an automatic mapping of the relationships only with the presentation of the input(s) and output(s) (procedure known as the learning process). The skill to deal with noisy data is also commented as an advantage compared with traditional regressive methods (Campbell (2006), Chen (2001)).

The disadvantages that can be described are related with the architecture chosen, normally large enough to accommodate a huge number of parameters that have to be predicted based on a small number of data points (whenever few historical data is available). Besides this shortcoming, it is often alleged that methods are not systematically tested and the results are not well commented and/or not presented in a clear way.

2.1. Artificial Neural Networks – Architecture, Activation Functions and Training Algorithms

The multilayer feed-forward architecture is still the preferred model in load forecasting applications, so it was also used in this work. Fig. 1 gives a simple example of a neural network with 2 neurons in the hidden layer to relate one output (y) with respect to three independent variables (x_1, x_2 and x_3). In order to avoid unnecessary heavy and complex structures, this report considers just one hidden layer of neurons between input and output layers. Different numbers of neurons in the hidden layer were carefully tested to compare different architectures, as it will be explained later in Section 4. The example of Fig. 1 is illustrated to exemplify the connections (with associated weights) between the elements (inputs/neurons and neurons/output). Each connection is simply characterized by a weight, for example, the connection between input x_1 to the second neuron is defined by the weight w_{12}.

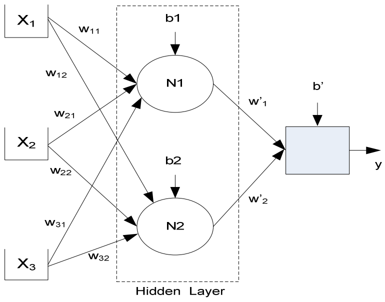

Figure 1 - Example of an ANN architecture with 2 neurons in the hidden layer

Each neuron, as the basic element of the architecture, processes the information captured in its inputs, such as the system's inputs or the previous hidden layers' outputs, by applying a function to the weighted sum of the inputs plus a scalar bias (Demuth (2007), Hagan (1996)).
Giving the example of the first neuron, it would then produce the following output:

$$\mathrm{output}(n_1) = f_1(x_1 \cdot w_{11} + x_2 \cdot w_{21} + x_3 \cdot w_{31} + b_1) \tag{1}$$

It must be noted that function f_1 is the activation function adopted. Typically the choice of activation functions depends on the forecasting case and the most popular functions are hyperbolic tangent function (*tansig*), sigmoid function and linear transfer function. The adoption of nonlinear functions is essential in cases of nonlinear relationships between outputs and inputs.

The output *y* is obtained applying the same methodology to the output layer:

$$y = f_2[\mathrm{output}(n_1) \cdot w'_1 + \mathrm{output}(n_2) \cdot w'_2 + b'] \tag{2}$$

The output activation function (f_2 in this case) can be distinct from activation functions applied in previous layers.
Historical data assumes an important role in the network training period, because it gives to the network the possibility to change its parameters (weights and biases) in order to minimize deviations between the real values and the estimations made. Several training methods are commonly used based on iterative optimizing methods. The objective is to minimize the performance function (typically based on the mean square error). One of the most popular training methods is the backpropagation algorithm, used to calculate the gradient of the network error (*E*) with respect to the network's modifiable weights and biases. Each weight and bias is updated using an iterative process of gradient descent (Hagan (1996)):

$$\Delta w_{ij} = -\varepsilon \cdot \frac{\partial E}{\partial w_{ij}} \tag{3}$$

$$\Delta b_j = -\varepsilon \cdot \frac{\partial E}{\partial b_j} \tag{4}$$

where ε is the learning rate used to be multiplied with the negative of the gradient to determine the changes in weights and biases. The learning rate controls the training speed. A higher rate would enable a faster algorithm but it may compromise convergence. On the other hand, small learning rates would make training considerably time-consuming. The training algorithm stops whenever one of the following conditions is verified (Demuth (2007)):

- a maximum number of epochs is attained;
- if the performance function is below a predetermined value;
- if the magnitude of the gradient is less than a predetermined value;
- a maximum training time was reached.

Another way to interrupt the training period is to use the cross-validation technique with the complete data set available divided into training sample and validation sample. The training sample is used to adapt weights and biases while the validation set is used to avoid overfitting (generalization inability). It is expected that during the training period, the error associated with the training sample is always being reduced but, whenever the error in the validation set starts to degrade, it means that the ANN becomes more accurate to characterize the training set, although, it doesn't represent an effective accuracy gain in characterizing new data (validation set).

A more recent training method is the Levenberg-Marquardt algorithm, an approximation to Netwon's method. As for the Newton method, a Jacobian matrix *J* must be obtained, containing the first derivatives of the performance function in respect of all network weights and biases. According to the Newton's method, the weights and biases update respects expression (5).

$$\Delta x = \left[J^T \cdot J\right]^{-1} \cdot J^T e(x) \tag{5}$$

with:

Δx – vector with network weights and biases updates;

J – Jacobian matrix;

e(x) – vector with output errors.

The Levenberg-Marquardt algorithm incorporates an adaptive parameter μ to maintain the search for good solutions in a faster way but also preserving convergence (Hagan (1996), Hagan (1994)).

$$\Delta x = \left[J^T \cdot J + \mu \cdot I\right]^{-1} \cdot J^T e(x) \tag{6}$$

with:

I – Identity matrix;

The adaptive parameter μ is multiplied by some factor β whenever a step results in an increased performance index. By the contrary, when a step reduces the performance function, the adaptive parameter μ is divided by β. With a large adaptive parameter, the algorithm becomes similar to a gradient descent algorithm with a small step size. Considering a reduced adaptive parameter, the algorithm tends to approximate the Newton's method. The algorithm therefore provides a compromise between the speed of Newton's method and the guaranteed convergence of steepest descent.

In this algorithm the specifications for the training period are the maximum number of epochs; a performance goal; a minimum performance gradient; the maximum number of validation failures; the maximum training time; the initial adaptive parameter μ; the associated μ increase and decrease factors as well as a maximum adaptive parameter.

2.2. Input variables usually adopted in neural networks for load forecasting

In the bibliography it is common to find some particular interest in the selection of the input vector in ANN architectures (Fidalgo (2005), Santos (2007)), using past load values identified by a justified entropy analysis (to discriminate the number of contiguous data values) and/or by a trend concept (to discriminate homologous periods, e.g. similar days of previous weeks) (Santos (2007)), or using exogenous factors such as macroeconomic variables, weather and seasonal variables. Recent approaches are also exploited to ascertain the most influent variables, e.g. using principal component analysis (Guo (2004)) to reduce the original input space to several characteristic variables or using support vector machines, an alternative to neural networks, that implements a structural risk minimization principle in regression thematic (Abbas (2006)).

There are some causes for unexpected and sudden load fluctuation, as the presence of special days like holidays (and normal days surrounding holidays) or recording failures. These anomalous load periods and some forms to minimize associated load forecasts deviations are well described in some works (Chicco (2001), Fidalgo (2007), Fidalgo (2005), Lamedica (1996)). The holidays can be treated imposing *a priori* that input patterns used to estimate the demand in those days must be affected by a reduction factor. Other suggestion is a binary variable inclusion in input vector, in order to give to the ANN a clear distinction between normal and special days. This effect brings two direct advantages: the ANN notes that holidays' forecasts are attenuated (reflecting less demand in these special days) and forecasts on normal days after the holidays should not depend blindly on holidays' data. References Chicco (2001) and Lamedica (1996) propose an unsupervised stage providing a classification of the historical data with Kohonen Self Organizing Maps, leading afterwards to separate learning processes of ANNs to forecast future demand. This supervised stage performs ANN training in data previously separated (typically putting in evidence the occurrence of special days) rather than on the entire data set.

The present work aims to explore the effect of integrate load profiles for different consumers classes in load forecasting issues and how this assumption can also be helpful to deal with difficulties triggered by special days.

3. Load Profiling

For the appropriate operation of competitive markets, the detailed knowledge of the hourly distribution of electric energy consumption is mandatory (the Portuguese legislation imposes a settlement period of 15 minutes (ERSE (2007))). The use of load profiles makes the settlement between distributors and suppliers possible, as the total energy record available in the traditional meters can be distributed through different interval periods. According to the new policies adopted by the energy markets, all customers should have access to the retail market, including small customers who are not subjected to hourly based load measurements. Load profiling is a way to obtain typical load shapes adjusted to specific groups of consumers, enabling their participation in the market. The elaboration of load profiles requires a monitoring stage for a representative sample of consumers, through a temporarily installation of hourly metering systems. The consequent classes' segmentation can be done following different strategies:

- Using clustering algorithms such as hierarchical clustering, fuzzy c-means or competitive neural networks is the most accurate approach. (Gerbec (2005)). The outputs are typical load profiles for each class, and the clustering algorithm can ensure homogeneous groups of consumers (in respect of energy distribution for a specific period) and also relevant differences between the groups formed. Several authors (Gerbec (2005), Matos (2005)) state that the use of clustering algorithms is the ideal way to segment ways of electric usage, however, a further nontrivial step of allocating profiles to new consumers is required, using specific attributes (such as billing data or activity sector information);

- Using predefined profiles for each activity sector (residential, commercial, industrial, services,…) (Jardini (2000));

- Using predefined load profiles with a segmentation based on billing information.

Following one of the segmentation techniques based on the energy use, the settlement between distributors and suppliers (being mandatory) becomes possible as the total energy record available in the traditional meters can be distributed through different interval periods.

In Portugal, load profiling has used a classification of Low Voltage (LV) consumers (with *rms* voltage between phases below or equal to 1 kV) based on contracted power and annual energy consumption (see Table I), respecting the principle of non-discrimination on the basis of energy use, imposed by the local regulatory authority (ERSE (2007)).

TABLE I
DESCRIPTION OF LOW VOLTAGE CLASSES
ADOPTED BY THE PORTUGUESE LEGISLATION

Class number	Class	Class Description
1	Special Low Voltage	Contracted Power > 41,4 kW
2	Normal Low Voltage – Class A	Contracted Power > 13,8 kVA but also Contracted Power ≤ 41,4 kVA
3	Normal Low Voltage – Class B	Contracted Power ≤ 13,8 kVA and Annual Consumption > 7140 kWh
4	Normal Low Voltage – Class C	Contracted Power ≤ 13,8 kVA and Annual Consumption ≤ 7140 kWh

Low Voltage – RMS Voltage between phases ≤1 kV

The load profiles for these low voltage classes can be examined in more detail in Fig. 2.

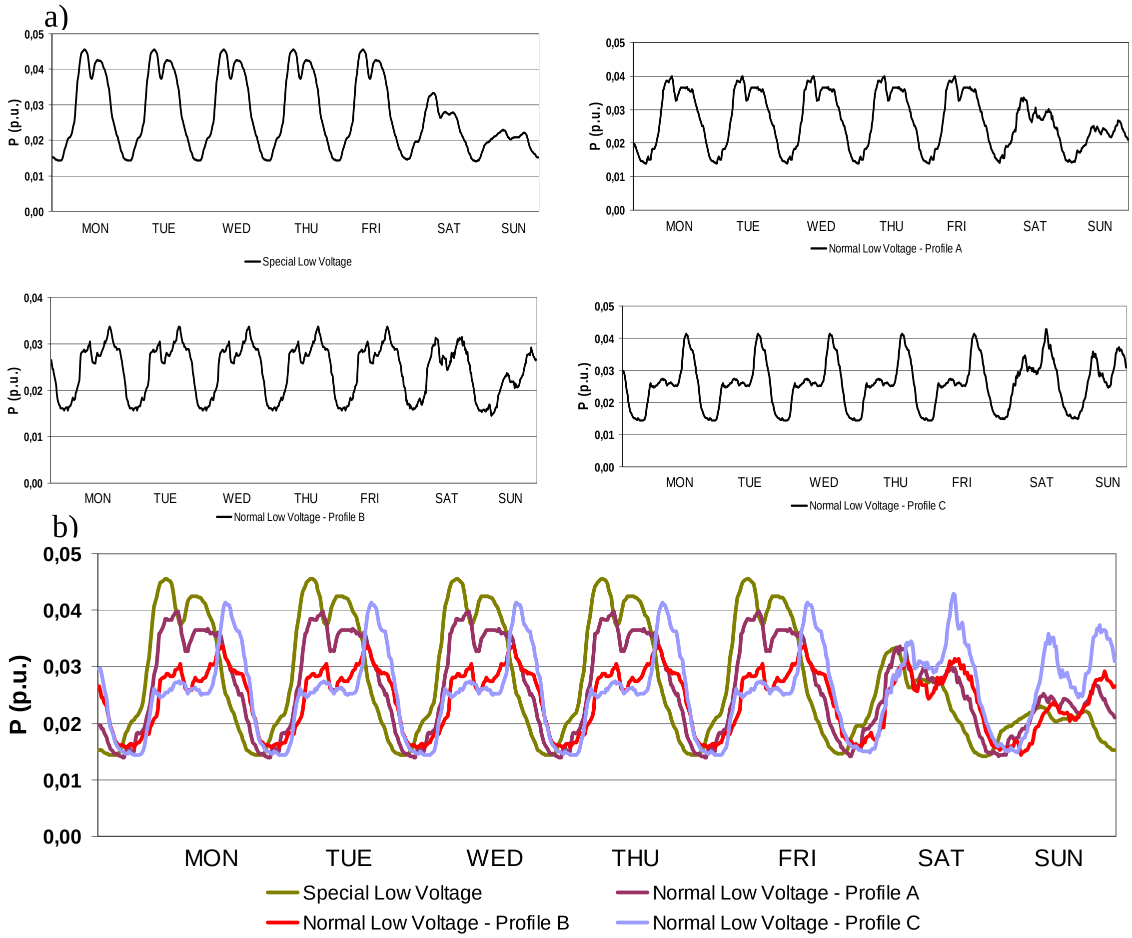

Figure 2: Low Voltage Load Profiles approved by the Portuguese Legislation (based on one complete week of October 2005) - a) Separated Load Profiles; b) Comparison between the Load Profiles adopted

The profiles are effectively set up in a monthly basis, so in a simple way, it is considered that during each month:

- all workdays assume a similar load shape;
- saturdays and sundays have particular profiles and also different from each other;
- in the presence of a holiday, the associated profile assumes a profile similar to a Sunday.

3.1. Integrating Load Profiling information in Load Forecasting Thematic

It was already assumed that load forecasting and load profiling are two subjects that have been treated independently, however from a manager point of view this separation can be irrelevant and even unnecessary (Espinoza (2005)). The load profiling is actually viewed as a way to enable the participation of the whole set of consumers in liberalized markets, and is particularly dedicated to small consumers that have not installed hourly metering systems. Load profiling simply used for billing purposes allow the settlement between the distribution operator and the suppliers in the market, based on the typical distribution of energy, given by load profiles, during the settlement period. Once the consumer’s usage of electricity is estimated in load profiles, this knowledge could be transposed to load forecasting issues (based on a bottom-up approach).

To investigate the load profile information effects in load forecasting, we used the load profiles approved by the Portuguese legislation, because this information is easily available, as it is made public by the independent regulator (before the inclusion of weather and seasonal effects – approved for the whole civil year) and by the transmission operator (after the inclusion of seasonal and weather effects).

As a first approach, simple assumptions were made regarding the segmentation, where only billing data were used, and regarding the different day types using a limited distinction only between workdays and weekend days. If the forecasting results become successful, it is expected that their accuracy may increase with the adoption of a better segmentation procedure and also with the bigger distinction between day types, using specific profiles for Saturdays, Sundays, Mondays and Fridays.

4. Case Study Description

This study uses consumption data collected from a sample of 1147 Low Voltage (LV) consumers with an integration period of 15 minutes. Besides this consumption data, a detailed commercial database is available for all consumers composing the sample. A random aggregation of a fixed number of 740 contributors’ consumers has led to an accumulated load diagram composition that was used to simulate a public substation typical load diagram. The electricity consumers monitoring was implemented in scope of a national project aiming to establish load profiles and create all necessary and mandatory conditions to allow a competitive market. These audits were available from 23-Mar-2003 to 23-Apr-2007, but the period effectively considered was from 16-Mar-2004 to 16-Mar-2006 (two complete years) because in this temporal horizon the consumption data was quite stable for the considered consumers without significant data error detections.

The consumption distribution through different classes was studied along the considered period.

For each hour, a description of the contributors' consumers was analysed and confronted with the commercial database where historical annual consumptions are available, and therefore, during the considered period, the total consumption can be distributed through different classes as seen in Fig. 3.

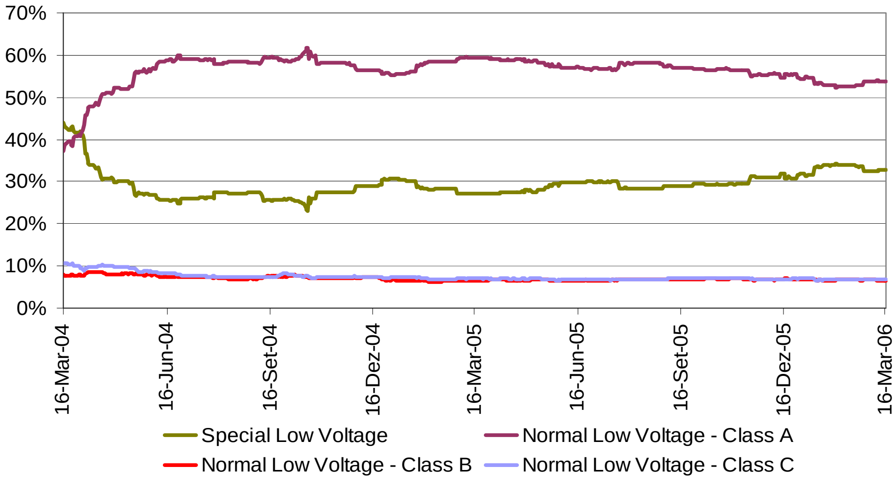

Figure 3 - Total consumption distributed through different classes for the considered period

The annual profiles available and used in the Portuguese market were explored to obtain a reconstructed load diagram for each class according to (7). It must be described that for Low Voltage consumers, the electricity market was only effectively opened in 2006, so for the three Low Voltage classes (Classes A, B and C) the load profiles considered during 2006 were transposed to the years 2004 and 2005, respecting the calendar shift (i.e. respecting the days of the week and the presence of holidays). Special low voltage load profiles were in fact approved for 2004, 2005 and 2006, so this information was effectively used in this study.

$$LD_{r\_class\,i\_hour\,h} = W_t \cdot CP_{class\,i\_hour\,h} \cdot \frac{LP_{class\,i\_hour\,h} \cdot CP_{class\,i\_hour\,h}}{\sum_{h}\sum_{i=1}^{4} LP_{class\,i\_hour\,h} \cdot CP_{class\,i\_hour\,h}} \tag{7}$$

$$i = 1,\ldots,4$$

where:

$LD_{r\_class\,i}$ is the reconstructed load diagram for class i;

W_t is the total consumption estimated for the aggregation (in a yearly basis);

$LP_{class\,i}$ is the load profile for class i;

$CP_{class\,i}$ is the consumption percentage of class i;

h is each considered hour [between hour 0 of 30-Mar-2004 until hour 23 of 16-Mar-2006].

The profiles contribution for each reconstructed load during one week can be exemplified in Fig. 4.

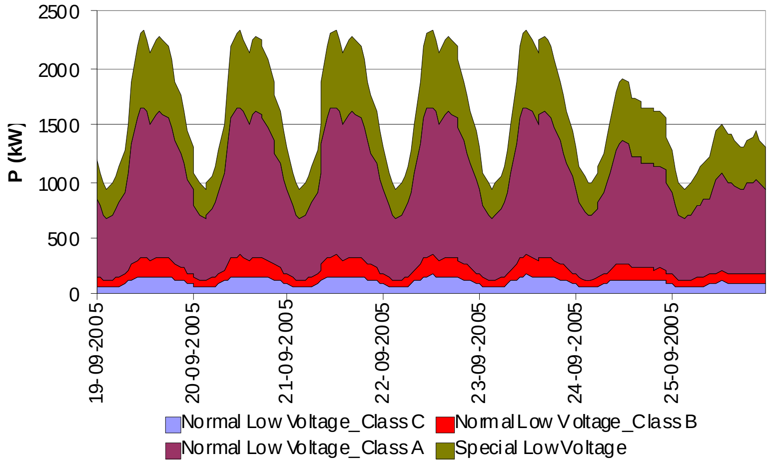

Figure 4: Reconstructed load diagrams obtained for each class – aggregation effect during one week of September 2005

In Fig. 5 it is possible to confront the aggregated reconstructed load diagram with the real one during one week.

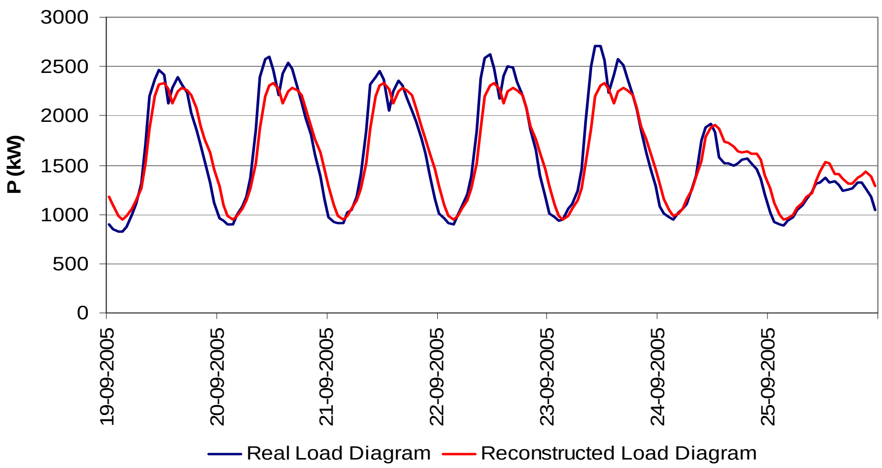

Figure 5: Comparison between real and aggregated reconstructed load diagrams during one week of September 2005

The assumption of similar workdays assumed by profiles leads to a considerable deviation from the real aggregated load diagram, as consumption patterns effectively vary according to the day of the week, mainly in non-domestic consumers. As the load profiles considered already include a treatment dedicated to the effect of special days (considering a load pattern similar to a Sunday), this information is in fact useful to approximate the reconstructed load diagram to the real one, as can be exemplified in Figure 6.

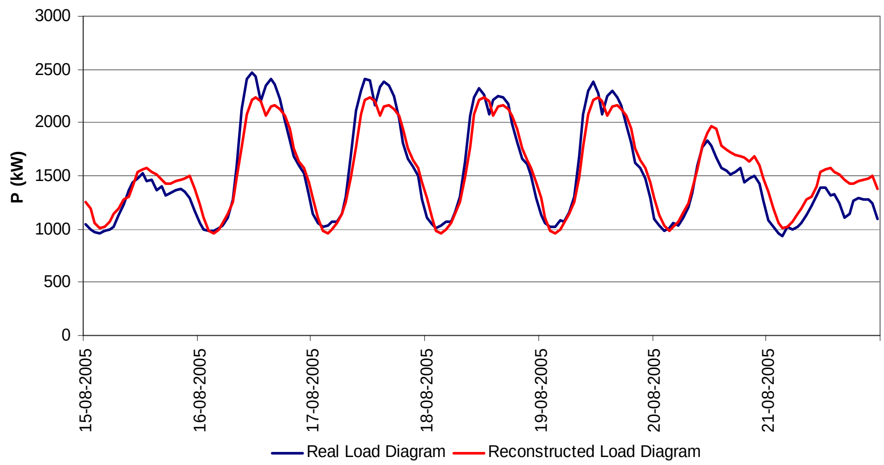

Figure 6: Comparison between real and aggregated reconstructed load diagrams during one week with a special day – 15th August is a religious holiday in Portugal

Despite some eventual inaccuracy associated with the aggregated reconstructed load diagram, this information was used as ANN inputs to forecast future load demand.

4.1. Models created in Load Forecasting using Artificial Neural Networks

The present work intends to evaluate and compare different models to forecast electric load for the 24 hours of the following day. The models were trained with data from 30th March 2004 to 31st December 2005, while the period from 1st January 2006 to 16th March 2006 was just used to test and compare the forecasting methodology.

The models were based on ANN structures following some specifications:

– all networks created do not have more than one hidden layer to avoid unnecessary complex models;

–all networks created use a *tansig* transfer function between input and hidden layers and a linear transfer function between hidden and output layers.

–all networks were trained using the Levenberg-Marquardt algorithm, knowing that it congregates reduced training time with demonstrated skills to guarantee convergence;

–for each model it was considered and evaluated a variable number of neurons in the hidden layer (from 3 to 15 neurons);

–for each model, and for different number of neurons in the hidden layer, twenty scenarios of network initialization were tested and compared (described in Fig. 7), bearing in mind that each different scenario creates distinct initial random network parameters (weights and biases);

–for these scenarios, it was also used a fixed and a random validation set in order to interrupt the training period in different ways (illustrated in Fig. 7). In both cases, data available from 30th March 2004 to 31st December 2005 were split in the following way: 80% for training data and 20% for validation data;

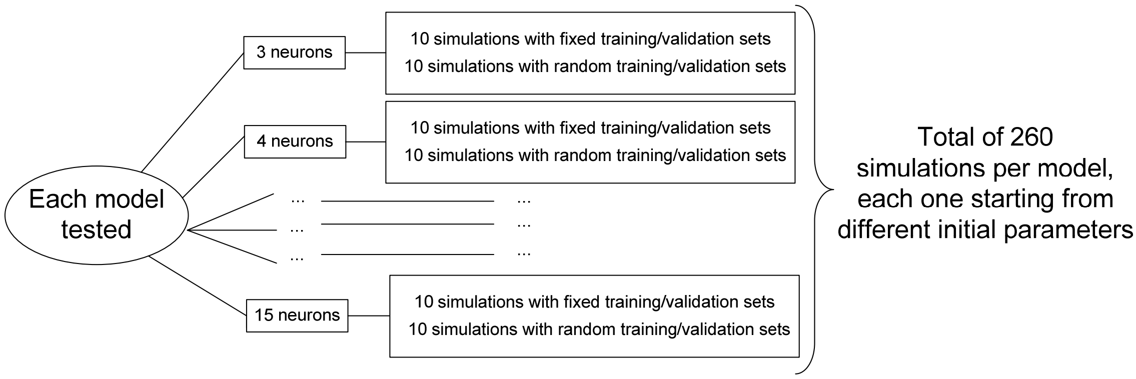

Figure 7: Detailed description of the simulation stage followed for each model

–network training parameters were defined to have a maximum of 1000 epochs, with a minimum gradient of 10^{-10}, an adaptive parameter μ initially adjusted to 0.001, a multiplier factor β set to 10/0.1 (to increase/or to decrease the adaptative parameter μ) and no training time limit. All simulations were effectively stopped by the cross-validation criterion.

Different models incorporating distinct input patterns were evaluated. The inputs and outputs were normalized in order to have zero mean and a unitary variance. In that way, the influence of different scales is attenuated. A description of models created is presented here:

Model I: this model is based on related works identified in the bibliography and uses 72 inputs such as hourly load values of the last day available, load values of the same weekday of previous week and load values of the same weekday two weeks before, considering these the most correlated with the variable(s) being predicted. For example, to forecast Thursday 15th September 2005 hourly load values (being estimated during Wednesday 14th September and with that day not entirely known), the model will use hourly load values of Tuesday 13th September 2005, Thursday 8th September 2005 and also Thursday 1st September 2005. This input vector decision was proved to be consistent after an autocorrelation analysis made to the time series available (see Fig. 8). Although the correlation with day *d-2* seems to be low, avoiding the influence of workdays in weekends and vice-versa, using a filter in order to just consider Wednesdays, Thursdays and Fridays (only dependent on workdays), makes the autocorrelation coefficient rise to 0.946, justifying the idea of incorporating the recent past.

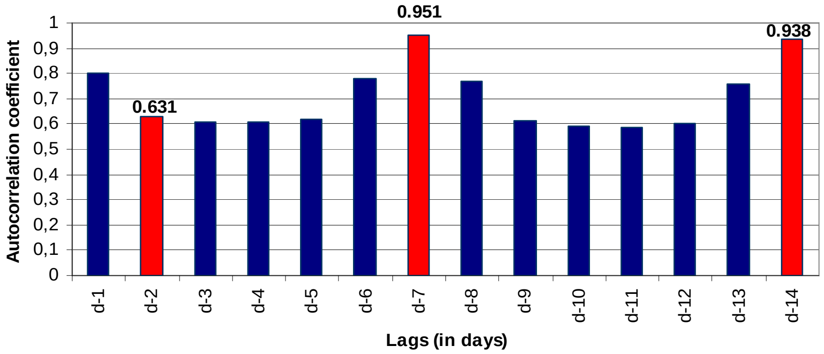

Figure 8: Autocorrelation coefficients identified for different lags in days

Model II: this model with 75 inputs incorporates three more inputs than Model I: one binary input to distinguish normal days and special days (like holidays) and two cyclical inputs to characterize day types.

Cyclical functions *sin* and *cos* were adopted to guarantee that each day has a univocal representation.

Model III: besides the 72 inputs of Model I, this perspective integrates the 24 hourly load values obtained from the reconstructed load diagram of the day being predicted. The correlation analysis between the real load diagram and the reconstructed load diagram also identified a Pearson correlation coefficient of about 0.937.

The comparison between several networks created for each model was possible using different error measures for the training sample and also for the testing sample. The error measures used in this study and normally used in similar works are:

– Mean Percentage Error (MPE):

$$\text{MPE }(\%) = \frac{\sum_h \left(\frac{\text{Real Load Value}_h - \text{Forecasted Load Value}_h}{\text{Real Load Value}_h} \right)}{\text{Number of hours considered}} \tag{8}$$

– Mean Absolute Percentage Error (MAPE):

$$\text{MAPE }(\%) = \frac{\sum_h \left(\frac{\left|\text{Real Load Value}_h - \text{Forecasted Load Value}_h\right|}{\text{Real Load Value}_h} \right)}{\text{Number of hours considered}} \tag{9}$$

– Root Mean Squared Percentage Error (RMSE):

$$\text{RMSE }(\%) = \sqrt{\frac{\sum_h \left(\frac{\text{Real Load Value}_h - \text{Forecasted Load Value}_h}{\text{Real Load Value}_h} \right)^2}{\text{Number of hours considered}}} \tag{10}$$

– Mean Absolute Deviation (MAD):

$$\mathrm{MAD}\ (\%) = \frac{\left(\dfrac{\sum_h \left| \mathrm{Real\ Load\ Value}_h - \mathrm{Forecasted\ Load\ Value}_h \right|}{\mathrm{Number\ of\ hours\ considered}} \right)}{\left(\dfrac{\sum_h \mathrm{Real\ Load\ Value}_h}{\mathrm{Number\ of\ hours\ considered}} \right)} \tag{11}$$

5. Results and Discussion

Some of the fundamental results to be presented are the error measures associated with different models analyzed. The errors were also detailed for the training period and the testing period, as appropriate neural networks must be able to generalize for data not used during the training stage. In the selection of the most suitable architecture for each model, it was evident that a reduced number of neurons in the hidden layer would lead to high training and testing errors, and also, that an excessive number of neurons can cause acceptable training errors but high testing errors. Table II gives detailed information about the error behaviour for the whole set of scenarios created, focusing in this case in the Mean Absolute Percentage Error (MAPE).

TABLE II

COMPARISON OF MAPE RESULTS INDICATORS OBTAINED FOR DIFFERENT SCENARIOS CREATED FOR THE THREE MODELS (FOR TRAINING AND TESTING DATA)

		Average MAPE	Standard Deviation in MAPE
Model I (best architecture with 7 neurons in the hidden layer)	Training Data	6.50%	1.03%
	Testing Data	9.70%	1.75%
Model II (best architecture with 9 neurons in the hidden layer)	Training Data	5.06%	1.01%
	Testing Data	7.63%	1.37%
Model III (best architecture with 3 neurons in the hidden layer)	Training Data	4.69%	0.67%
	Testing Data	7.38%	1.8%

With Model II, the average MAPE value is smaller than with Model I, and Model III gets the best result with an average of 4.69% for training data and 7.38% for testing data. The analysis of the MAPE standard deviation shows that, despite some scenarios with eventual inappropriate network initializations or even with a validation set that could interrupt the training too early, thus compromising the overall forecasting accuracy, forecasting errors in Model II are more concentrated and close to the average than errors in Model I. It was with the Model III that the best average results were obtained, but, also the bigger standard deviation (testing data). The use of a larger number of inputs by Model III makes it more sensitive to the quality of network initializations.

In order to provide the most successful forecasts, the results shown below are exclusively related with the best scenario found for each model from several network simulations applied.

TABLE III

COMPARISON BETWEEN DIFFERENT MODELS FOR THE BEST SCENARIOS DIFFERENT ERROR MEASURES FOR TRAINING AND TESTING DATA

	Error measures	MPE	MAPE	RMSE	MAD
Model I	Training Data	-0.76%	5.60%	9.00%	5.60%
	Testing Data	1.93%	7.09%	11.47%	7.25%
Model II	Training Data	-0.74%	4.94%	6.74%	4.96%
	Testing Data	0.71%	5.54%	7.26%	5.83%
Model III	Training Data	-0.36%	4.26%	5.69%	4.09%
	Testing Data	1.95%	4.96%	6.76%	5.08%

Analyzing the results of the three models in Table III for both the training and testing data, it can be shown a considerable reduction in forecasting errors, making evident that the adoption of Model III turns the forecasting procedure even more accurate.

A special attention was dedicated to the forecasting problems caused by special days. Bearing in mind that the load profiles already include information about these special days with the subsequent consumption reduction effect (probably with some simple assumptions because holidays profiles are assumed similar as Sundays profiles), this information can thus be passed to the load forecasting process. Fig. 9 exposes the comparison of different error measures associated with special days for the models analyzed.

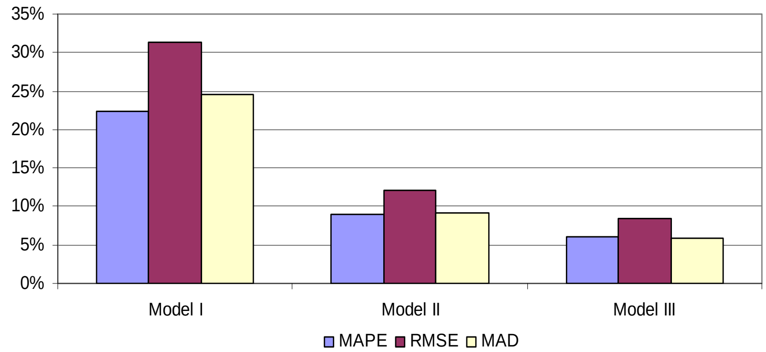

Figure 9: Detailed analysis of Error Measures identified in different models for special days (holidays) of training set

Model I still reveals serious limitations to forecast future load demand, being particularly evident in special days. For instance, RMSE measure presents an unacceptable result, demonstrating that the model is strongly dependent of large individual errors normally verified in special days. Model II reduces these errors because it includes a binary variable to distinguish special days and normal days, so the neural network becomes more sensitive to these special days during the training process. Besides using historical data, Model III also integrates future load demand prediction (available by the use of load profiles for different classes) and it demonstrates important benefits to the forecasting methodology. To enhance the model viability, some comparisons in forecasting specific special days are available in Figs. 10 and 11.

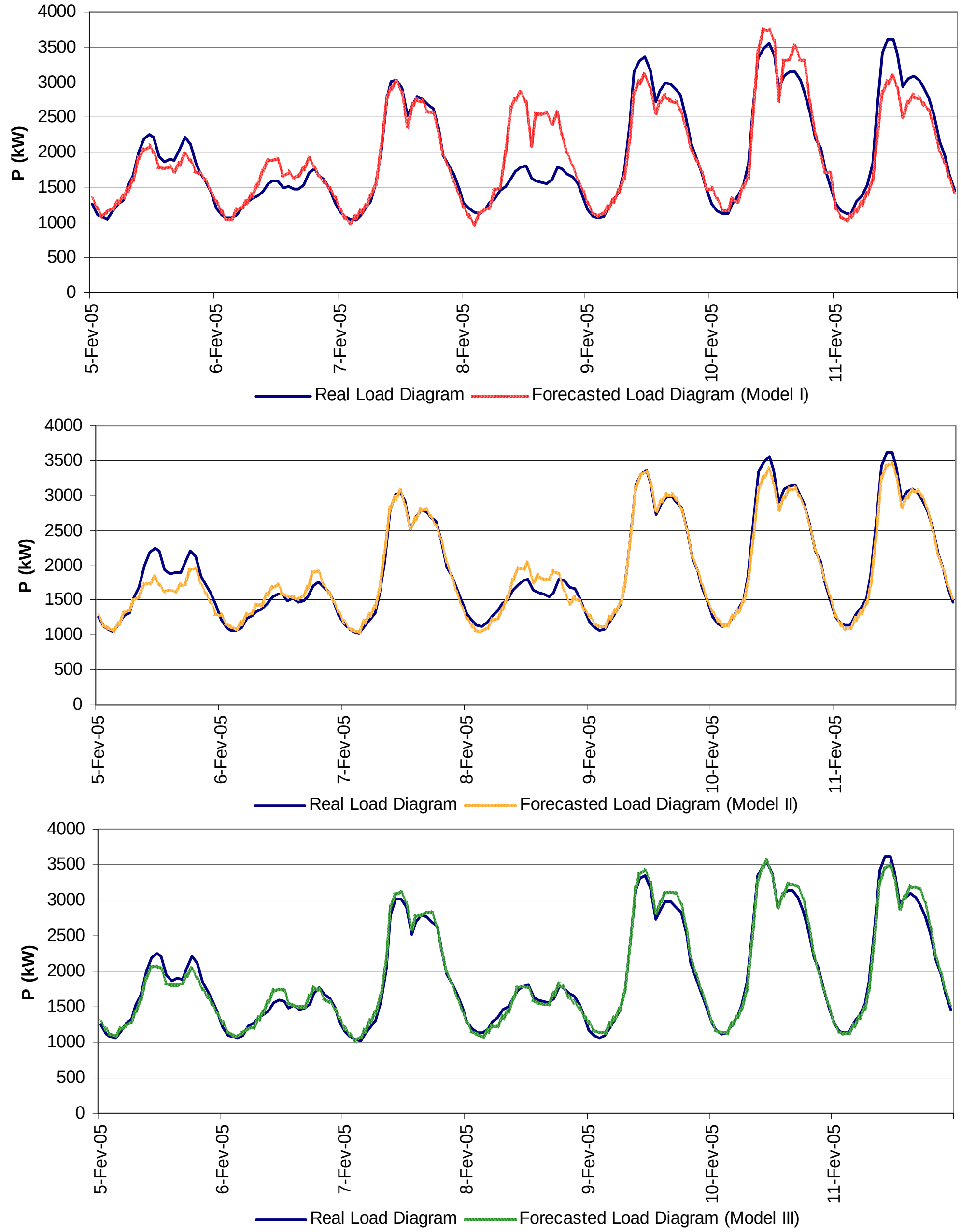

Figure 10: Comparison between models to forecast one week with a special day - Tuesday, 8th February 2005 corresponds to the carnival day

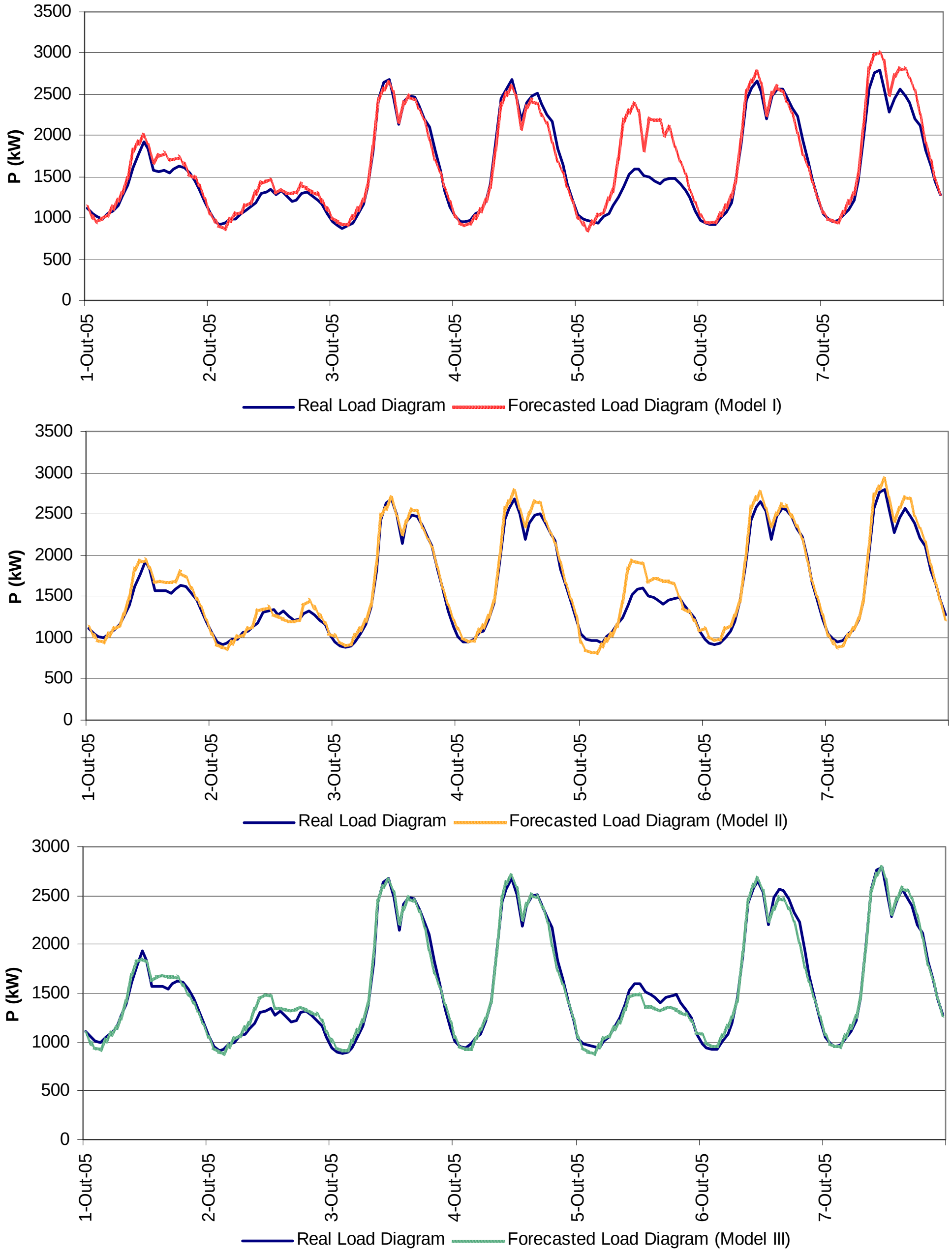

Figure 11: Comparison between models to forecast one week with a special day - Wednesday, 5th October 2005 corresponds to a national holiday in Portugal

Although Model III is the one with more inputs, it doesn't mean that it should need more neurons in the hidden layer to model possible relationships between outputs and inputs. In fact, comparing the proposed models discussed in this Section, it can be said that Model III is the one with less neurons in the hidden layer. Model I uses 7 neurons, Model II was trained with 9 neurons, while the Model III analyzed only uses 3 neurons. In the hidden layer of an ANN architecture used in Model III, more than 5 neurons only serves to degrade forecasting errors in the testing period, which means that the ANN can be appropriate to the training sample but does not generalize this capability to the testing set. A brief description of the neural network used in Model III is shown in Fig. 12.

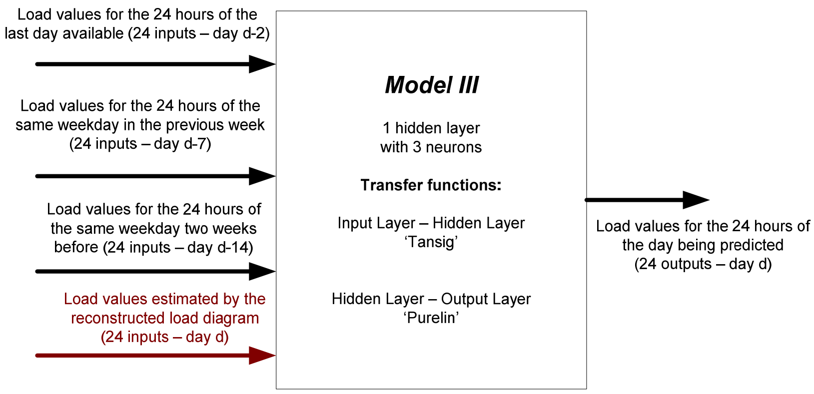

Figure 12: ANN Architecture for Model III

6. Conclusions and Further Work

This work intends to analyse the effect of integrating information obtained from available load profiles to load forecasting techniques based on neural networks. The detailed consumers' behaviour knowledge could be, in that way, used to forecast future global load demand. As this case study has some particularities such as the global load diagram, which was obtained from a LV consumers' aggregation, it is not our intention to compare this case study with other load forecasting works already discussed and applied to real distribution or public substations. The load profiling methodology adopted is also a questionable approximation, knowing that classes' segmentation based on clustering methods could be more accurate. Nevertheless, even with those assumptions, the results presented show that the method represents a promising technique in load forecasting. It is recognized that the viability of the technique strongly depends on adequate consumer segmentation, load profiles accuracy representing each class, proper selection of neural network input vector, lead time to predict (one hour, one day, one week,…) and updated commercial information to characterize the individual contributions of a global load diagram.

In order to get better results in further work and also to make the methodology applicable in realistic cases, it can be described some aspects need to be improved, namely:

– the adoption of a more accurate methodology to segment consumers;

– the creation of distinct load profiles for different day types (evaluate the need to deal with more than one profile for workdays and dedicate more attention to holidays effects – distinction between fixed and mobile holidays);

– the analysis of the artificial neural networks sensitivities, evaluating the effect of the different inputs in each output. In that way it can be discriminated the importance of the vector associated with the reconstructed load diagram among the remaining inputs;

– the adoption of an incremental training during the testing phase, making possible the network adaptation to seasonal or unexpected load changes.

7. References

Abbas, S. R., and Arif, M. (2006), "Electric Load Forecasting Using Support Vector Machines Optimized by Genetic Algorithm", *INMIC'06 IEEE Multitopic Conference 2006*, 23-24 Dec. 2006, pp. 395-399

Campbell, P. and Adamson, K. (2006), "Methodologies for Load Forecasting", *3rd International IEEE Conference on Intelligent Systems*, September 2006, pp. 800-806

Chen, H., Canizares, C.A. and Singh, A. (2001), "ANN-based short-term load forecasting in electricity markets", *IEEE Power Engineering Society Winter Meeting*, 28th January2001 – 2nd February 2003 , vol. 2, pp. 411-415

Chicco, G., Napoli, R. and Piglione, F. (2001), "Load pattern clustering for short-term load forecasting of anomalous days", *Proceedings of the IEEE PowerTech 2001*, Porto, 10-13 Sept 2001, Vol. 2, pp. 6

Demuth, H., Beale, M. and Hagan, M. (2007), "Neural Network Toolbox 5 User's Guide", Revised for Version 5.1, September 2007, Copyright 2005-2007 by the Mathworks Inc.

ERSE – Energy Services Regulatory Authority (2007) , "Guide of Data Measuring, Reading and Availability", document nº 4591-A/2007

Espinoza, M., Joye, C., Belmans, R. and DeMoor, B. (2005), "Short-Term Load Forecasting, Profile Identification, and Customer Segmentation: A Methodology Based on Periodic Time Series", *IEEE Transactions on Power Systems*, August 2005, Volume: 20, Issue: 3, pp. 1622-1630

Fidalgo, J. and Matos, M. A. (2007), "Forecasting Portugal Global Load with Artificial Neural Networks", *ICANN2007 - International Congress on Artificial Neural Networks 2007*, pp. 728-737

Fidalgo, J. N. and Peças Lopes, J. A. (2005), "Load Forecasting Performance Enhancement When Facing Anomalous Events", *IEEE Transactions on Power Systems*, February 2005, Vol. 20, No. 1, pp. 408-415

Gerbec, D., Gasperic, S., Smon, I. and Gubina, F. (2005) , "Allocation of the Load Profiles to Consumers using Probabilistic Neural Networks", *IEEE Transactions on Power Systems*, May 2005, Vol. 20, No. 2, pp. 548- 555

Guo, X., Chen, Z., Ge, H. and Liang, Y. (2004), "Short-term load forecasting using neural network with principal component analysis", Proceedings of the *International Conference on Machine Learning and Cybernetics*, 26-29 Aug. 2004, Vol. 6, pp. 3365-3369

Hagan, M. T., Demuth, H. B. and Beale, M. (1996), “Neural network design”, PWS Publishing Company, Boston MA.
Hagan, M. T. and Menhaj, M. B. (1994), “Training Feedforward Networks with the Marquardt Algorithm”, *IEEE Transactions on Neural Networks*, Vol. 5, No. 6, November 1994, pp. 989-993

Hippert, H., Pedreira, C. and Souza, R. C. (2001), “Neural Networks for Short-Term Load Forecasting: A Review and Evaluation”, *IEEE Transactions on Power Systems*, February 2001, Vol. 16, No. 1, pp. 44-55

Jardini, J., Tahan, C., Ahn, S. and Figueiredo, F. M. (2000), “Daily Load Profiles for Residential, Commercial and Industrial Low Voltages Consumers”, *IEEE Transactions on Power Delivery*, January 2000, Vol. 15, No. 1, pp. 375-380

Lamedica, R., Prudenzi, A., Sforna, M., Caciotta, M. and Cencellli, V. (1996), “A neural network based technique for short-term forecasting of anomalous load periods”, *IEEE Transactions on Power Systems*, Nov. 1996, Volume: 11, Issue: 4, pp. 1749 – 1756

Matos, M., Fidalgo, J., and Ribeiro, L. (2005), “Deriving LV Load Diagrams for Market Purposes using Commercial Information”, *Proceedings of the 13th International Conference on Intelligent Systems Application to Power Systems*, 6-10 November 2005, pp. 105-110

Santos, P., Martins A. and Pires A. (2007), “Designing the input vector to ANN-based models for short-term load forecast in electricity distribution systems”, *International Journal of Electrical Power and Energy Systems*, May 2007, Vol. 29 - Issue 4, pp. 338-347

Willis, H. L. (2004), “Power Distribution Planning Reference Book”, Marcel Dekker, Inc., 2004, Second Edition, Revised and Expanded