

Article

Artificial Neural Network for Short-Term Load Forecasting in Distribution Systems

Luis Hernández ^{1,*}, Carlos Baladrón ², Javier M. Aguiar ², Lorena Calavia ², Belén Carro ², Antonio Sánchez-Esguevillas ², Francisco Pérez ³, Ángel Fernández ⁴ and Jaime Lloret ⁵

¹ Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT—Research Center for Energy, Environment and Technology), Autovía de Navarra A15, Salida 56, 42290 Lubia, Soria, Spain

² Departamento Teoría de la Señal y Comunicaciones e Ingeniería Telemática, Universidad de Valladolid, E.T.S.I. Telecomunicación, Campus Miguel Delibes, Paseo de Belén 15, 47011 Valladolid, Spain; E-Mails: cbalzor@ribera.tel.uva.es (C.B.); javagu@tel.uva.es (J.M.A.); lcaldom@ribera.tel.uva.es (L.C.); belcar@tel.uva.es (B.C.); antsan@tel.uva.es (A.S.-E.)

³ Departamento Ingeniería Mecánica, Universidad de Zaragoza, Escuela de Ingeniería y Arquitectura, Calle María de Luna 5, 50018 Zaragoza, Spain; E-Mail: franjope@hotmail.com

⁴ Departamento de Tecnología Electronica, Universidad Rey Juan Carlos, Escuela Superior de Ciencias Experimentales y Tecnología, Calle Tulipán s/n, 28933 Móstoles, Madrid, Spain; E-Mail: angelfernandezgago@hotmail.com

⁵ Departamento de Comunicaciones, Universidad Politécnica de Valencia, Camino Vera s/n, 46022 Valencia, Spain; E-Mail: jlloret@dcom.upv.es

* Author to whom correspondence should be addressed; E-Mail: luis.hernandez@ciemat.es; Tel.: +34-97-5281013; Fax: +34-97-5281051.

Received: 7 February 2014; in revised form: 7 March 2014 / Accepted: 10 March 2014 /

Published: 17 March 2014

Abstract: The new paradigms and latest developments in the Electrical Grid are based on the introduction of distributed intelligence at several stages of its physical layer, giving birth to concepts such as *Smart Grids*, *Virtual Power Plants*, *microgrids*, *Smart Buildings* and *Smart Environments*. *Distributed Generation (DG)* is a philosophy in which energy is no longer produced exclusively in huge centralized plants, but also in smaller premises which take advantage of local conditions in order to minimize transmission losses and optimize production and consumption. This represents a new opportunity for renewable energy, because small elements such as solar panels and wind turbines are expected to be scattered along the grid, feeding local installations or selling energy to the grid depending

on their local generation/consumption conditions. The introduction of these highly dynamic elements will lead to a substantial change in the curves of demanded energy. The aim of this paper is to apply *Short-Term Load Forecasting (STLF)* in *microgrid* environments with curves and similar behaviours, using two different data sets: the first one packing electricity consumption information during four years and six months in a *microgrid* along with calendar data, while the second one will be just four months of the previous parameters along with the solar radiation from the site. For the first set of data different *STLF* models will be discussed, studying the effect of each variable, in order to identify the best one. That model will be employed with the second set of data, in order to make a comparison with a new model that takes into account the solar radiation, since the photovoltaic installations of the *microgrid* will cause the power demand to fluctuate depending on the solar radiation.

Keywords: microgrid; short-term load forecasting; multi-layer perceptron; artificial neural network

1. Introduction

Since the origins of the electrical system, engineers have always tried to understand its operation, which was relatively simple then, but has become so difficult to manage and control in our days. For example, Spencer and Hazen [1] present in 1925 an artificial system to simulate and study Kirchhoff's laws with different loads, and with alternating and direct current.

Hamilton [2] makes a study of the addition of curves for characterization of the load curve on the basis of a set of meaningful features: peak load, minimum load and load factor; posing also the prediction of the load as a challenge for the operation. Forrest [3] presents the statistical treatment of historical data, the study of the load fluctuation with the climatic conditions and its forecasting, as fundamental to avoid breakage of electrical installations because of climatic anomalies. Along with the above, Davies [4] employs linear regression to analyze the relationship between the electrical demand and the weather, and its application in the load estimate. As a curiosity, Gruetter [5] uses punch cards from public administration for the prediction of the load, based on statistical averages of consumption behaviours.

Therefore, the electrical system has always been in search of better forecasting tools for the demand. For the distributors, the main objective of the prediction has been to find the energy necessary to better adjust their offers in the daily and intra-day markets. The work of Matthewman and Nicholson [6] gave us, already in 1968, indications about the most interesting time horizons for demand prediction, speaking about *Short-Term Load Forecasting (STLF)*. However, as it is described by Hippert *et al.* [7], the prediction of the electricity demand can be categorized (among other criteria) depending on the time interval to predict, what is commonly known as horizon of prediction:

- *Very Short-Term Load Forecasting (VSTLF)*: in the very short term, from a few seconds or minutes to several hours.
- *STLF*: from a few hours to weeks.

- *Medium/Long-Term Load Forecasting (M/LTLF)*: in the medium and long term, from months to years.

The most important horizons of prediction are weekly, daily and hourly. Having a good prediction of the following 24 h plays a vital role for companies, because that can influence the optimal functioning of the hourly planning of the generation units, purchase/sale in exchange systems.

As Hernández *et al.* [8] show, since the 1970s until now, different architectures have been employed to solve the problem of the predictions of electricity demand using models based on linear and non-linear techniques. In addition, many hybrid models have been used in combination of some of the above. Those based on *Artificial Neural Networks (ANNs)* stand out within the non-linear. They may be of the supervised type *Multi-Layer Perception (MLP)* (Garcia-Ascanio and Matt [9]), not supervised as *Self-Organizing Map (SOM)* (Marin *et al.* [10]), or combination of them in several stages (Hernández *et al.* [11]). As [8] present, the models based on *ANNs* require a less intellectual effort on the part of researchers, since they do not need to obtain the complicated linear equations that attempt to relate the potential nonlinearities associated with the problem of demand prediction; the complexity of the models based on *ANNs* is in its optimization, with the necessity of having tests that vary the internal topology of the network (number of neurons in the hidden layer and learning function).

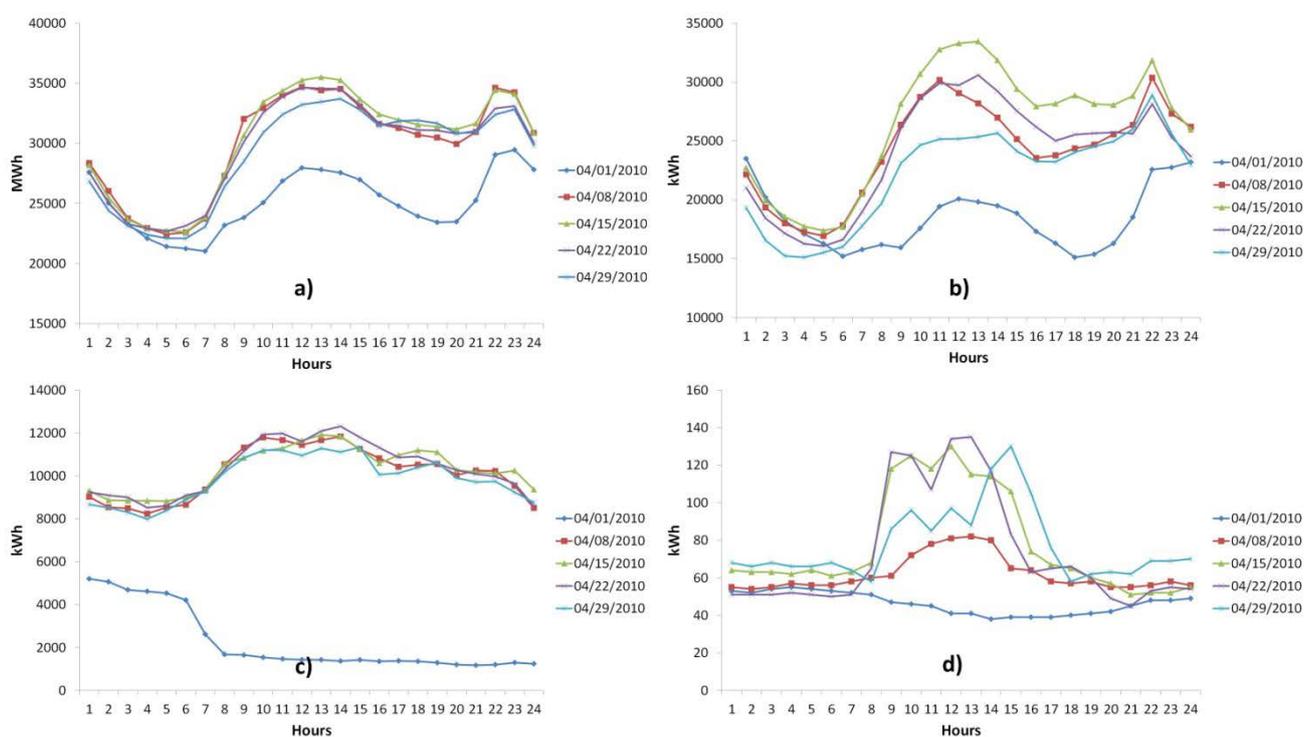
Continuing with [8], they show the need to keep on working with the predictive models for demand and generation, especially in the new environments characterized by incorporating distributed intelligence in its physical layer, as can be the *Smart Grids (SGs)*, *Virtual Power Plants (VPPs)*, *microgrids*, *Smart Buildings* and *Smart Environment. Distributed Generation (DG)* represents a new opportunity for the sources of renewable energy generation, but referred to as small power plants and in the sites where there is consumption. This *DG*, in the final consumption, will lead to a substantial change in the load curves of consumers that install it, and to have a good vision of the new energy performance, it will be vital to take into account the climatic variables from which new applications and business models will emerge (Hernández *et al.* [12]). In addition, the use of intelligence and *Smart Meters (SMs)* in the new environments, will provide new data that reflect more accurately the behaviour of the consumption point or consumption/production (building, neighborhood, industrial estate, town, city, *etc.*); electricity is clearly dependent on external factors (social, economic, *etc.*), therefore, the provision of new data will allow you to have a more updated behaviour profile.

Hagan and Behr [13] (as early as 1987), emphasized on the limitations of linear models compared to the non-linear techniques based on *ANNs*. In addition, Badri [14] found extremely interesting models based on *ANNs* when you have short series of data (a limitation which is commonly found in the new environments *Smart Environments—SG, VPP, microgrid* and *Smart Building—*which shall have recent data through *SMs*). By completing the above, in [8] it was concluded that these new environments will be provided with models, whose accuracy will depend on the availability of measures of key variables that make them efficient, such as local climatic variables, behaviour and state of users, early programming of the operation of their consumption points, *etc.* This availability of recent information along with the easiness of incorporation of new variables in the models will make the use of models based on *ANN* more attractive.

With regard to load curves, there are important differences between curves of aggregate environments (countries, regions, *etc.*) and curves of disaggregated environments (*SG, VPP, microgrid*

and *Smart Building*), since in the former, load curves will present a smooth and less variable profile and therefore are more easily predictable, while in disaggregated environments load curves will present dynamic behaviours, with abrupt and random changes in their profiles which make their prediction difficult (Hernández *et al.* [15]). In Figure 1 load curves of four environments of very different scales are represented: country, small city (of a size similar to a *microgrid*), industrial estate (*microgrid* potential) and a *microgrid* (experimental environment); all the load curves are for the Thursdays of April, 2010, and the 04/01/2010 corresponds to a holiday in all the sites. The diverse scales can be seen in the different representations in Figure 1, according with the consumption of the four sites.

Figure 1. (a) Country; (b) Small City (potential *microgrid*); (c) Industrial Park (potential *microgrid*); (d) *Microgrid*.



A common feature in all of them is that the bank holidays have a totally different behaviour from the working days, but in addition, there are marked differences when comparing the four sites, as well as if the load curves of the rest of days (non-festive days) are observed. It can be seen that the profiles of the load curves in Figure 1a are smoother than the profiles of the other sites, and as the site becomes smaller, the load curves present more abrupt profiles. The load curves of (c) are completely different (in terms of shape) with regard to (a) and (b), this is due to the fact that as it is an industrial park, it presents a consumption behaviour which is very dependent on the hours when the productions of the manufacturing plants are concentrated, and/or the working hours of the service companies; this information will be essential for the aggregators, as it is shown in [11,16]. A greater difference is shown in (d), not only comparing the load curves with (a–c), but also if one compares the load curves between them, as there is a type of *microgrids* where their consumption patterns depend on the operation of the major points of consumption that compose them, and the situation of the *DG* installed

and its type. This variability in the profiles of the load curves that appear in these environments will make the selection of the input variables fundamental which make up the models for the prediction of the demand.

SGs and *microgrids* are undergoing a drastic change in order to deal with increasingly diversified and various service requests. Zhou *et al.* [17] propose a QoE-driven power scheduling in the context of *SG* from the perspectives of architecture, strategy and methodology. Zhou and Rodrigues [18] present an integrated and efficient middleware for heterogeneous services in *SG*; specifically, they first develop a mutual application access control principle that keeps users obtaining a satisfying assignment; next, a collaborative and dynamic information exchange infrastructure is proposed for different *SMs*, and a local information collector is designed to implement the communication and computing through power management messages.

The aim of this paper is to apply *STLF* in *microgrid* environments with curves and similar behaviours as in Figure 1d. In these environments the operation of their points of consumption is crucial to shape the aggregated load curve, and therefore, the randomness of its profile increases with respect to other larger environments. Two different datasets will be employed for validation of the *STLF* models: the first one will comprise four years and six months of electricity consumption data in the *microgrid* along with calendar data (working days, day of the week, *etc.*), while the second one will pack just four months of the same parameters, but including also solar radiation in-site. For the first set of data different models will be discussed, adding new variables, to make *STLF* and select the best model. The best model will be used to make *STLF* with the second set of data, and so it's possible to compare it with a new model that takes into account the solar radiation, since the photovoltaic installations of the *microgrid* will cause the power demand to be higher or lower based on the solar radiation.

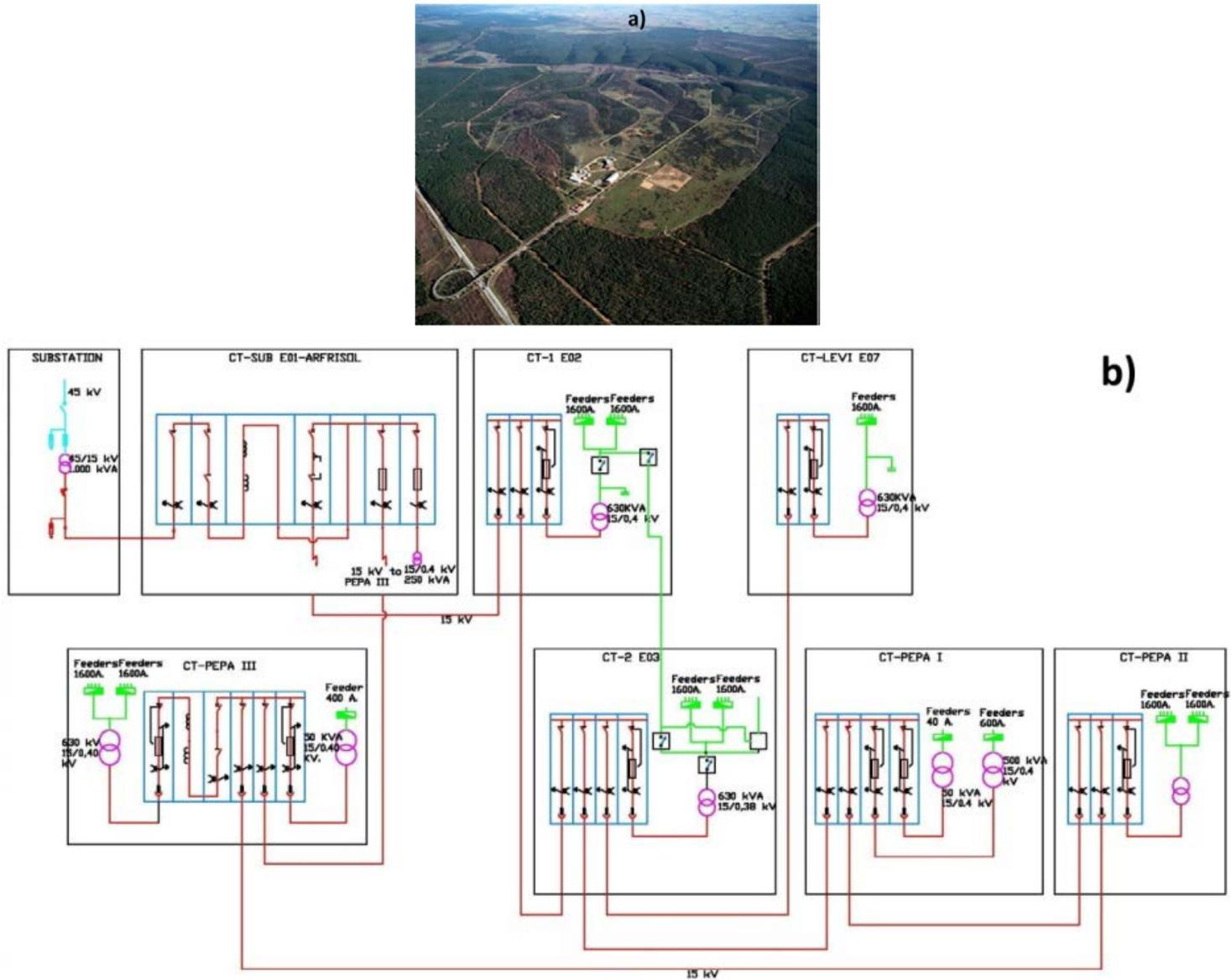
The paper is organized as follows: Section 2 shows the data used as well as the methodology used and the architecture models. Section 3 shows the indicators for the performance assessment. Section 4 will present the obtained results. In Section 5 the results obtained in the previous section will be analyzed. Finally, in Section 6 the conclusions will be presented and future works will be defined.

2. Data Description and Methodology Framework to Be Used in the Study

2.1. Microgrid CEDER-CIEMAT

The *Centre for the Development of Renewable Energy Sources (CEDER)* is located only a few kilometres from the city of Soria, in Spain. It belongs to the Public Research Organization *CIEMAT*, attached to the Ministry of Economy and Competitiveness. *CEDER-CIEMAT* is presented as an environment of similar characteristics to a *microgrid*, ideal for demonstration and experimentation applied to elements of *DG*, storage, power electronics and distributed intelligence, from laboratory testing to testing in a real space. Figure 2a shows a portion of the 640 hectares of *CEDER-CIEMAT*, where the project *MIREDCON* among others is to be carried out.

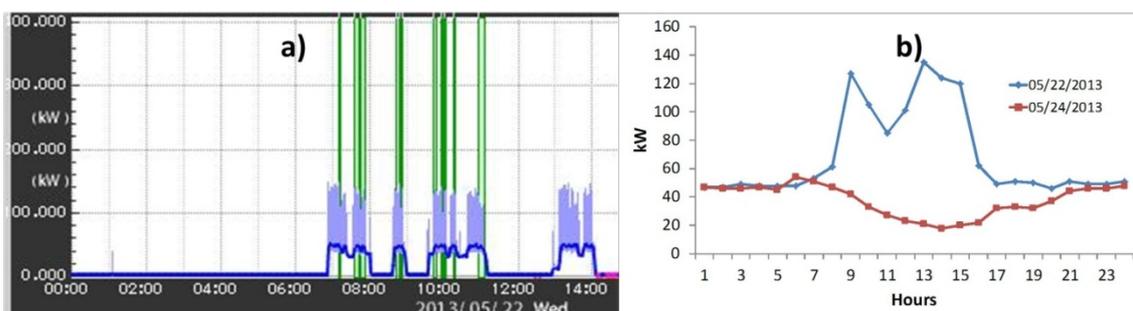
Figure 2. (a) Aerial photography of CEDER-CIEMAT; (b) Medium voltage electrical diagram of the microgrid.



The utility produces an electrical supply at 45 kV, and through a transformer at the entrance of CEDER-CIEMAT it is converted to 15 kV. The medium voltage is divided by the entire site by means of seven centers of transformation, as shown in the diagram in Figure 2b. The low voltage which supplies the diverse consumption, mainly the researchers' buildings and trial plants (manufacture of pellets and combustion boilers) parts from the different processing centres. The manufacture of pellets is performed in two industrial buildings; the process of milling (three different mills) is carried out in one of them, while in the second one the process of pelletization (a single plant) takes place. The combustion tests are performed in the factory premises, in which there are five combustion boilers of different powers.

The operation of the pelletization and combustion test plants is considered fundamental to outline the load curves of the *microgrid*. Figure 3a shows the power profile of the manufacture of pellets when one of the windmills of pelletization is active; as it can be noticed, the electricity consumption is adjusted to the hours in which the plant is operating, which clearly has an impact in the final load curve of the *microgrid*, as it can be seen in the blue curve in Figure 3b. Therefore, in this type of site the operation of their plants will cause the load curve to be in one way or another, as it can be seen when comparing the blue curve (plant operation) with the red curve (without plant operation) in Figure 3b, and for that reason greater randomness is added on the profiles of the load curves.

Figure 3. Load Curve (kW): (a) electrical consumption of pelletization mill on 05/22/2013; (b) total electricity consumption of the *microgrid*, blue curve day 05/22/2013 and red curve 05/24/2013.



In the *microgrid* and on a decentralized basis, there are three mini-photovoltaic (PV) plants installed and producing electrical power (45 kW electrical output at inverters). The PV production considered as a single producer, has already been considered in the past under the name of *Virtual Power Plant (VPP)*, with its consequent benefits from the economic, operation and control point of view (Wille-Haussman *et al.* [19], Hernández *et al.* [20]). In the case of the *microgrids*, and when it is not available the production measure, the PV production will be taken into account in the case of *STLF*, because if the percentage of PV production in the *microgrid* is high compared to its potential consumption, the days with high solar radiation will impact in less need of electricity from the utility. In the case of CEDER-CIEMAT, the percentages of PV production compared to the total consumption of the *microgrid*, varies in average 2% on weekdays and 20% on weekends, so that the same day of a month with a similar operation of plants than another may have an overall consumption of the *microgrid* far superior as you have a lower solar radiation than the counterpart.

2.2. Load Data

In Section 2.1 the elements (consumption and generators) available in the *microgrid* have been detailed. However, the only data available to be able to make *STLF* are data from the SM located at the point of connection with the utility, therefore, there is no visibility of PV generation. These data, together with those presented in Sections 2.2.1 and 2.2.2 will be those used for the different models. For this paper, two different sets of data have been used, because data from solar radiation are not available to the entire volume of existing data in the Data Set A, therefore, we have had to opt to have a second Data Set B to be able to use the solar radiation. The two sets of data (A and B) will be explained below, and will be used for two different experiments, which will be explained in Sections 2.2.1 and 2.2.2.

2.2.1. Data Set A

This data set will be employed at experiment A. The load curves of the SM of *CEDER-CIEMAT* (point of connection to the utility) are available from 1 January 2009 to 30 June 2013. The value of the energy consumption will be noted each time, being the units kWh. Load curves are complemented by the day of the week, the month and working/non-working days. In the case of the day of the week it has been opted for the following encoding: *Sunday* = 0, *Monday* = 1, *Tuesday* = 2, *Wednesday* = 3, *Thursday* = 4, *Friday* = 5 and *Saturday* = 6. In regard to working/non-working days, the codification will be: *day's work* = 1 and *holiday* = 2. For months, the codification will be: *January* = 1, *February* = 2, *March* = 3, *April* = 4, *May* = 5, *June* = 6, *July* = 7, *August* = 8, *September* = 9, *October* = 10, *November* = 11 and *December* = 12.

In the models presented in Section 3, the variables that have cyclical nature (day of the week and month) will be introduced as two variables (instead of one) in the form of sine and cosine. It has been demonstrated that this transformation improves the performance of the ANN (Drezga *et al.* [21]). In the case of the day of the week and the month, each variable entered in the form of sines and cosines improves prediction (Hernández *et al.* [15], Ramezani *et al.* [22], Razavi and Tolson [23]).

2.2.2. Data Set B

From the end of January 2013, *CEDER-CIEMAT* has a *Baseline Surface Radiation Network (BSRN)* in its *microgrid*, so solar radiation data will be available. In these types of problems, we use the global solar radiation variable, not using the rest of possible variables (ultraviolet, infrared, *etc.*); special attention should be paid to the hour, as the datalogger records solar time, a treatment of solar data is required to put the records in solar time. Therefore, there is a set of data from February 2013 until May 2013 formed by the same fields as the exposed in Section 2.2.1 plus solar radiation.

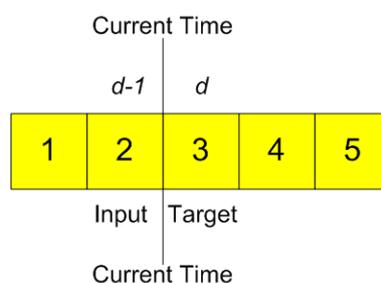
The best model obtained from the raised models set forth in Section 2.3.1 will be used to predict a volume of days again, lower than the one which will be obtained from the Data Set A, but this time with the Data Set B. The results will be compared with a new model, based on the winner, but to which solar radiation will be added as a new input variable.

2.3. Methodology and Structure

Below the different prediction models applied to each set of data will be explained. In the same way, the paper will discuss the script used to optimize the model (number of neurons in the hidden layer and function of learning). For all the models the architecture is based on *Multi-Layer Perceptrons (MLP)*, which will require a learning function on a set of patterns, and subsequently will require validation in a phase of operation with another set of data not used in the learning.

Independent of the Data Set (A or B), and for all models, the total of the available data is divided as follows: 70% for the learning phase and 30% for the operation phase. The philosophy of the architecture raised both for the learning of the *MLP* and the phase of operation, which is where the prediction will be done for the next day, follows the model shown in Figure 4. Suppose you want to predict day 3 (d), when day 2 ($d - 1$) has finished and the data for that day are available, it will be time to make the prediction of day 3 (d), and so on. The above explanation is also valid for the manner in which the patterns will be presented during the learning phase. Independent of the foregoing, some of the input variables may be the estimate of a value of d , as you'll see in the following sections.

Figure 4. Philosophy raised for the learning phase and for the operation phase.



The way to optimize the *MLPs*—both to determine the number of neurons in the hidden layer and to establish the best learning algorithm—is usually performed by a heuristic method. Therefore, an automatic script developed in MatLab will be used, where all possible parameters (number of neurons in the hidden layer, learning function, network performance function during learning, *etc.*) will be varied to select the best topology for each of the proposed networks.

For each of the models, for each of the learning functions, and for each hidden layer size (between 10 and 30 neurons in the experiment with Data Set A and between 1 and 30 neurons in the experiment with Data Set B), the script executed 100 different runs in order to statistically achieve meaningful results which rule out the random factors influencing the *ANN* (such as the initial state). The reason for the change in the number of neurons from 1 to 30 in the models of the experiment with Data Set B is that the number of neurons in the hidden layer clearly depends on the number of input patterns, number of variables in the input layer and number of variables in the output layer; in the models with Data Set B, the number of patterns in the learning phase will be 84, compared to the 1132 present when you use the Data Set A.

The learning functions considered are: *traingd* is gradient descent backpropagation; *traingdm* is gradient descent with momentum backpropagation; *traingda* is gradient descent with adaptive learning rate backpropagation; *traingdx* is gradient descent with momentum and adaptive learning rate backpropagation; *trainrp* is resilient backpropagation; *traincgf* is conjugate gradient backpropagation

with Fletcher-Reeves updates; *traincgp* is conjugate gradient backpropagation with Fletcher-Ribière updates; *traincgb* is conjugate gradient backpropagation with Powell-Beale restarts; *trainscg* is scaled conjugated gradient backpropagation; *trainbfg* is BFGS quasi-Newton backpropagation; *trainoss* is one-step secant backpropagation; *trainlm* is Levenberg-Marquardt backpropagation; and *trainbr* is Bayesian regulation backpropagation. The number of neurons in the hidden layer is detailed in the next section of this paper, while the optimal network performance function is *Sum Squared Error (SSE)*.

2.3.1. Model Structures of Experiment with Data Set A

Below the posed models will be presented, which will be trained and validated with the Data Set A. For all models, the output variables (prediction) will be the 24 values of consumption of the day to predict. The models are the following:

- *Forecast with Load (F_L)*: the input variables of this model will correspond with the 24 values of consumption of the day preceding the prediction;
- *Forecast with Load and Working/non-Working (F_L_W)*: The Working/non-Working variable is added on that day and the day to predict to the variables mentioned in the model *F_L*. The codification of Working/non-Working was commented on in Section 2.2.1;
- *Forecast with Load, Working/non-Working and Day of the week (F_L_W_DW)*: The day of the week of the day to predict and the previous one is added to the variables mentioned in the *F_L_W* model. The codification of the day of the week was commented on in Section 2.2.1;
- *Forecast with Load, Working/non-Working and Month (F_L_W_M)*: The month of the day preceding the prediction is added to the variables mentioned in the *F_L_W* model. The codification of the month was commented on in Section 2.2.1.

The final structure of the model *F_L* will be the following:

Input: 24

- $L_{1(d-1)}, L_{2(d-1)}, \dots, L_{23(d-1)}, L_{24(d-1)}$: 24 values of consumption hours of the day $d - 1$.

Output: 24

- $L_{1(d)}, L_{2(d)}, \dots, L_{23(d)}, L_{24(d)}$: 24 values of consumption hours of the day to predict d .

Hidden: the hidden layer will depend on the results of the script; the details will be seen in Section 4.

The final structure of the model *F_L_W* will be the following:

Input: 26

- $L_{1(d-1)}, L_{2(d-1)}, \dots, L_{23(d-1)}, L_{24(d-1)}$: 24 values of consumption hours of the day $d - 1$;
- $W_{(d-1)}, W_{(d)}$: working/non-working of the day $d - 1$ and the day d .

Output: 24

- $L_{1(d)}, L_{2(d)}, \dots, L_{23(d)}, L_{24(d)}$: 24 values of consumption hours of the day to predict d .

Hidden: the hidden layer will depend on the results of the script; the details will be seen in Section 4.

The final structure of the model $F_L_W_DW$ will be the following:

Input: 30

- $L_{1(d-1)}, L_{2(d-1)}, \dots, L_{23(d-1)}, L_{24(d-1)}$: 24 values of consumption hours of the day $d - 1$;
- $W_{(d-1)}, W_{(d)}$: working/non-working of the day $d - 1$ and the day d ;
- $DW_{(d-1)}, DW_{(d)}$: day of the week of $d - 1$ and d . Each one of them is entered as two, in the form of sine and cosine.

Output: 24

- $L_{1(d)}, L_{2(d)}, \dots, L_{23(d)}, L_{24(d)}$: 24 values of consumption hours of the day to predict d .

Hidden: the hidden layer will depend on the results of the script; the details will be seen in Section 4.

The final structure of the model $F_L_W_M$ will be the following:

Input: 30

- $L_{1(d-1)}, L_{2(d-1)}, \dots, L_{23(d-1)}, L_{24(d-1)}$: 24 values of consumption hours of the day $d - 1$;
- $W_{(d-1)}, W_{(d)}$: working/non-working of the day $d - 1$ and the day d ;
- $M_{(d-1)}, M_{(d)}$: month of $d - 1$ and d . Each one of them is entered as two, in the form of sine and cosine.

Output: 24

- $L_{1(d)}, L_{2(d)}, \dots, L_{23(d)}, L_{24(d)}$: 24 values of consumption hours of the day to predict d .

Hidden: the hidden layer will depend on the results of the script; the details will be seen in Section 4.

2.3.2. Model Structures of Experiment with Data Set B

The model with better performance during the operation phase of the ones set forth in Section 2.3.1 (it is $F_L_W_M$) will be re-tested with the Data Set B. The aim will be to compare it with a new model, which will be explained below. The new model will be the following:

- *Forecast with Load, Working/non-Working and Solar Radiation ($F_L_W_M_SR$)*: accumulated solar radiation of the day to predict and the previous day is added to the model $F_L_W_M$.

The final structure of the model $F_L_W_M_SR$ will be the following:

Input: 32

- $L_{1(d-1)}, L_{2(d-1)}, \dots, L_{23(d-1)}, L_{24(d-1)}$: 24 values of consumption hours of the day $d - 1$;
- $W_{(d-1)}, W_{(d)}$: working/non-working of the day $d - 1$ and the day d ;
- $M_{(d-1)}, M_{(d)}$: month of $d - 1$ and d . Each one of them is entered as two, in the form of sine and cosine;
- $SR_{(d-1)}, SR_{(d)}$: accumulated solar radiation of the day $d - 1$ and the day d .

Output: 24

- $L_{1(d)}, L_{2(d)}, \dots, L_{23(d)}, L_{24(d)}$: 24 values of consumption hours of the day to predict d .

Hidden: The hidden layer will depend on the results of the script; the details will be seen in Section 4.

3. Evaluating the Performance of ANN

For all the validation phase, a daily *MAPE* is estimated by means of:

$$MAPE_d = 100 \cdot \frac{\sum_{i=1}^{24} \left| \frac{L_{(d)i} - \hat{L}_{(d)i}}{L_{(d)i}} \right|}{24} \quad (1)$$

where $L_d(i)$ represents the measured value; $\hat{L}_d(i)$ represents the estimated value and i the hour in both cases.

The mean error along all samples in the validation phase is estimated by means of:

$$ERROR_{OP} = \frac{\sum_{i=1}^n MAPE_{di}}{n} \quad (2)$$

where n represents the number of days of the validation phase.

To examine how the forecasting error is reflected on the load curve, error is displayed on a graphic including all forecasted days in the validation phase. The forecast mean error for each of the 24 h is obtained by means of:

$$MAPE_h_i = 100 \cdot \frac{\sum_{k=1}^n MAPE_{i,k}}{n} \quad (3)$$

with $i = 1, 2, \dots, 24$; n stands for the sample size in the validation phase and $MAPE_{i,k}$ the hourly error i for the day k .

For each number of neurons in the hidden layer 100 iterations are executed. A new mean is necessary here to obtain the hourly *MAPE* taking into account the number of iterations considered:

$$MAPE_iter_i = \frac{\sum_{k=1}^{iter} MAPE_hi_k}{iter} \quad (4)$$

where $k = 1, 2, \dots, 100$ represents each iteration and $iter$ the number of iterations used (100) and $i = 1, 2, \dots, 24$.

Finally, the total *MAPE* used to compare the solution for each number of neurons in the hidden layer (and to choose the optimum number) is calculated as follows, obtaining the mean value of the 24 h considered:

$$MAPE_total = \frac{\sum_{i=1}^{24} MAPE_iter_i}{24} \quad (5)$$

where $i = 1, 2, \dots, 24$ represents each hour.

In order to develop an analysis of the results, a monthly *MAPE* will be necessary:

$$MAPE_month_k = \frac{\sum_{k=1}^m MAPE_d_k}{m} \quad (6)$$

where $k = 1, 2, \dots, m$ represents each day; j the considered month and m the number of days contained in the considered month.

4. Results

Table 1 shows the *MAPE* (5) for the validation phase of each of the four models proposed in Section 2.3.1 (Data Set A) to forecast the 24 values of the next day's consumption. For each model, all the learning functions and the optimal number of neurons are shown, for all days in the validation phase. The optimal values (the ones that minimize *MAPE* during validation phase) are shown in bold. Similarly, Table 2 (the models of Table 2 need a smaller number of neurons than models in Table 1) shows the *MAPE* (5) for the validation phase of each of the four models proposed in Section 2.3.2 (Data Set B).

The best model of Data Set A is *F_L_W_M*, with a *MAPE* of 15.34% for the learning function *trainbr* and 21 neurons in the hidden layer of *MLP*. The best model of Data Set B is again *F_L_W_M*, with a *MAPE* of 16.69% for the learning function *trainbr* and 16 neurons in the hidden layer of *MLP*, compared to a *MAPE* of 18.40% for the model *F_L_W_M_SR* (*trainbr* and 26 neurons).

The above is due to the fact that Data Set B uses less patterns in the learning phase than Data Set A, in line with Bishop [24], who introduced the concept of “*curse of dimensionality*”, which says that the number of patterns to do mapping grows exponentially with the dimension of the input space, so networks with a high input space will require a great amount of patterns for the learning.

Figure 5 shows the evolution of *MAPE* (5) during the validation phase depending on the learning functions and number of neurons for all the models of the Data Set A and with 100 iterations in each case. Similarly, Figure 6 shows the evolution of *MAPE* (5) for all the models of the Data Set B.

Table 1. Comparison of the minimum *MAPE* (5) for all the forecast models proposed (Experiment with Data Set A) and for all learning functions. Where: (1) optimal number of neurons for the specific model and learning functions; (2) *MAPE* (%) for the optimal number of neurons; and (3) standard deviation (%) for the optimal number of neurons.

Models	<i>traingd</i>			<i>traingdm</i>			<i>traingda</i>			<i>traingdx</i>			<i>trainrp</i>			<i>traingcf</i>			<i>traingcp</i>		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
F_L	11	203.37	141.04	10	206.40	146.56	14	30.19	6.60	20	26.54	8.82	14	22.27	1.61	18	22.35	1.88	19	22.13	1.77
F_L_W	10	207.67	127.69	11	207.78	125.82	15	25.39	5.48	10	207.67	127.69	12	17.06	1.41	30	16.65	1.30	21	16.31	1.33
F_L_W_DW	10	207.62	127.45	11	207.71	125.80	15	25.28	5.30	11	198.10	123.70	10	16.78	1.25	29	16.67	1.39	30	16.64	1.31
F_L_W_M	10	187.92	107.44	10	178.12	107.43	11	24.87	6.70	12	20.10	4.37	14	16.66	1.53	20	16.43	1.32	13	16.40	1.18

Models	<i>traingcb</i>			<i>traingcg</i>			<i>traingbf</i>			<i>traingss</i>			<i>trainglm</i>			<i>traingbr</i>		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
F_L	19	22.40	1.88	20	22.38	1.90	13	25.08	2.84	20	22.73	1.84	10	22.82	2.78	10	24.22	1.49
F_L_W	22	17.05	1.32	20	16.59	1.35	13	19.94	2.53	20	17.23	1.28	12	16.66	1.52	11	17.10	0.82
F_L_W_DW	21	16.62	1.41	25	16.89	1.60	13	20.23	2.88	30	17.29	1.36	10	16.89	2.34	10	16.87	1.00
F_L_W_M	28	16.43	1.37	16	16.40	1.27	11	19.56	2.29	25	16.46	1.41	16	17.59	2.20	21	15.34	0.85

Table 2. Comparison of the minimum *MAPE* (5) for all the forecast models proposed (Experiment with Data Set B) and for all learning functions. Where: (1) optimal number of neurons for the specific model and learning functions; (2) *MAPE* (%) for the optimal number of neurons; and (3) standard deviation (%) for the optimal number of neurons.

Models	<i>traingd</i>			<i>traingdm</i>			<i>traingda</i>			<i>traingdx</i>			<i>trainrp</i>			<i>traingcf</i>			<i>traingcp</i>		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
F_L_W_M	1	42.94	20.16	1	42.32	18.56	5	20.80	6.68	6	21.32	6.94	4	19.39	5.42	2	19.99	5.62	1	19.38	3.22
F_L_W_M_SR	1	42.58	18.70	1	44.20	18.64	4	21.71	7.70	3	22.09	7.84	3	21.93	5.83	4	21.79	6.60	1	19.75	3.56

Models	<i>traingcb</i>			<i>traingcg</i>			<i>traingbf</i>			<i>traingss</i>			<i>trainglm</i>			<i>traingbr</i>		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
F_L_W_M	1	20.33	6.06	7	18.87	4.84	1	19.42	3.82	1	19.53	3.34	1	19.53	3.38	16	16.69	3.17
F_L_W_M_SR	1	21.34	3.73	3	20.97	6.04	1	19.94	2.65	1	20.58	6.01	1	22.50	5.26	26	18.40	4.21

Figure 5. MAPE (%) (5) in each of the models for the Data Set A: (a) F_L ; (b) F_L _W; (c) F_L _W_DW; (d) F_L _W_M.

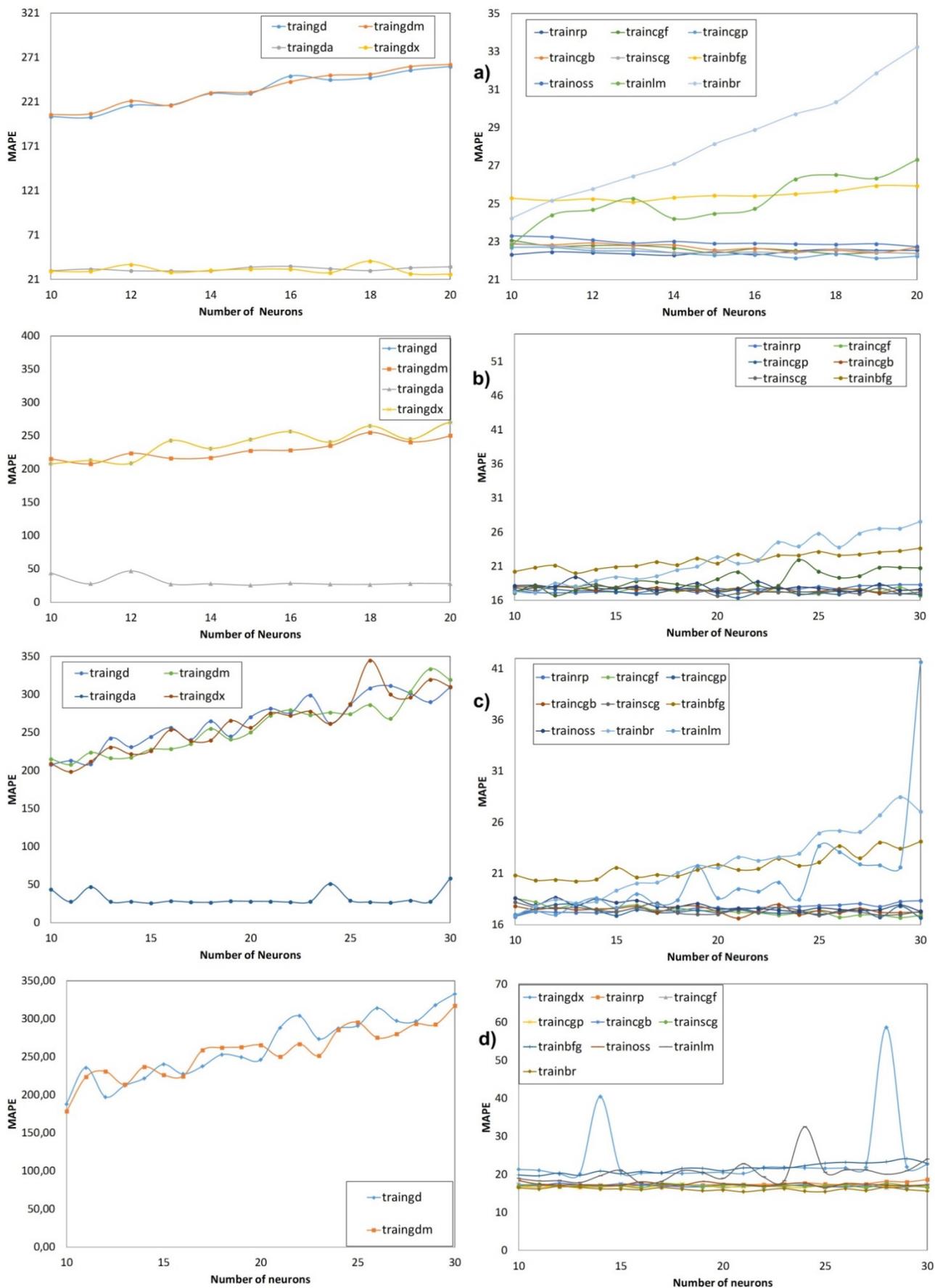
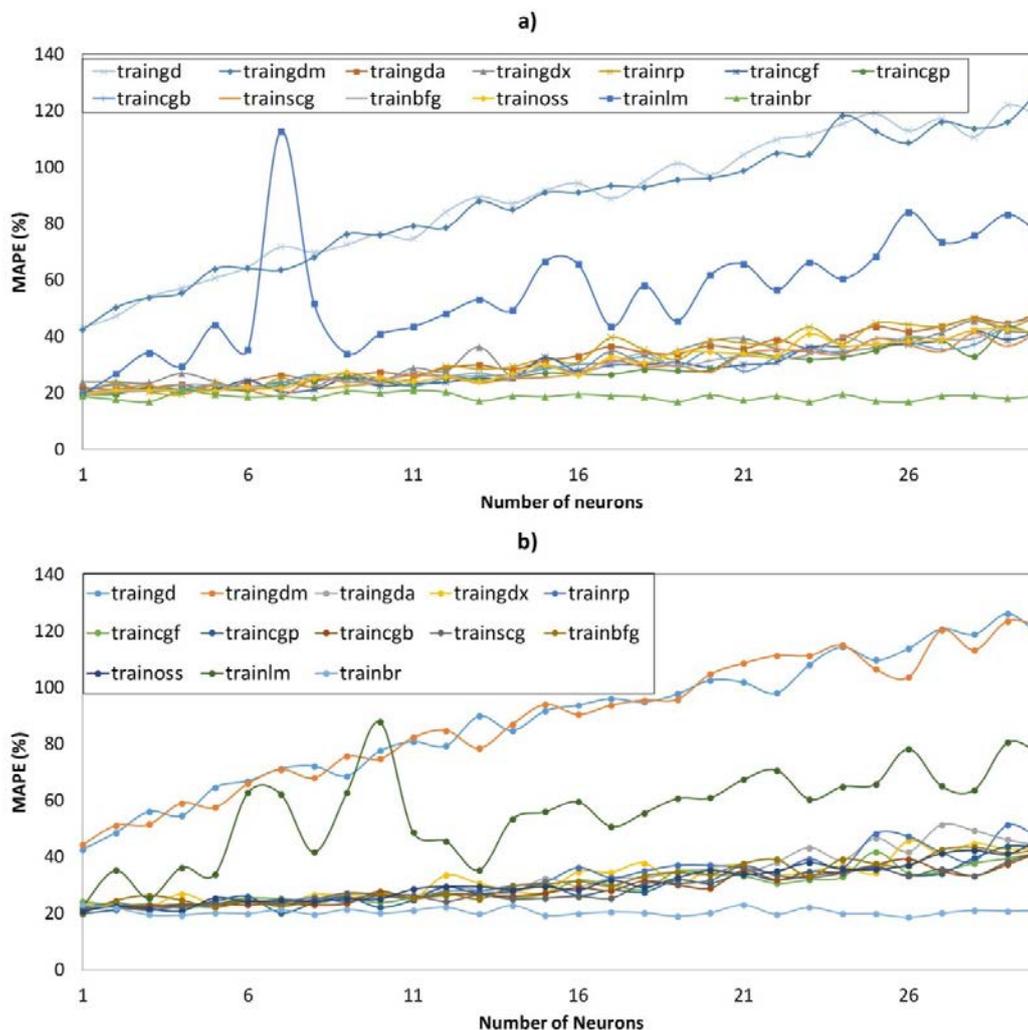


Figure 6. *MAPE* (%) (5) in each of the models for the Data Set B: (a) *F_L_W_M*; (b) *F_L_W_M_SR*.



5. Result Analysis

This section presents the analysis of results. The analysis is presented for the Data Set A (Section 5.1) and the Data Set B (Section 5.2).

The winning architecture that is mentioned several times in this section corresponds with the results of one of the iterations executed for the optimum number of neurons in the hidden layer.

5.1. Result Analysis for Data Set A

Figure 7 shows the hourly average *MAPE* (3) of the winning architecture (all models). The best one is in line with the results of Table 1, where the lowest *MAPE* corresponds to *F_L_W_M* model (*F_L_W_M* is the best model). In all models, the highest errors occur in the central hours (9–14 h); these hours coincide with the operation of the plants.

Figure 7. Hourly Average *MAPE* (%) (3) of the winning architecture (all models): (a) *F_L* (19 neurons, *traincgp*); (b) *F_L_W* (21 neurons, *traincgp*); (c) *F_L_W_DW* (21 neurons, *traincgb*); (d) *F_L_W_M* (21 neurons, *trainbr*).

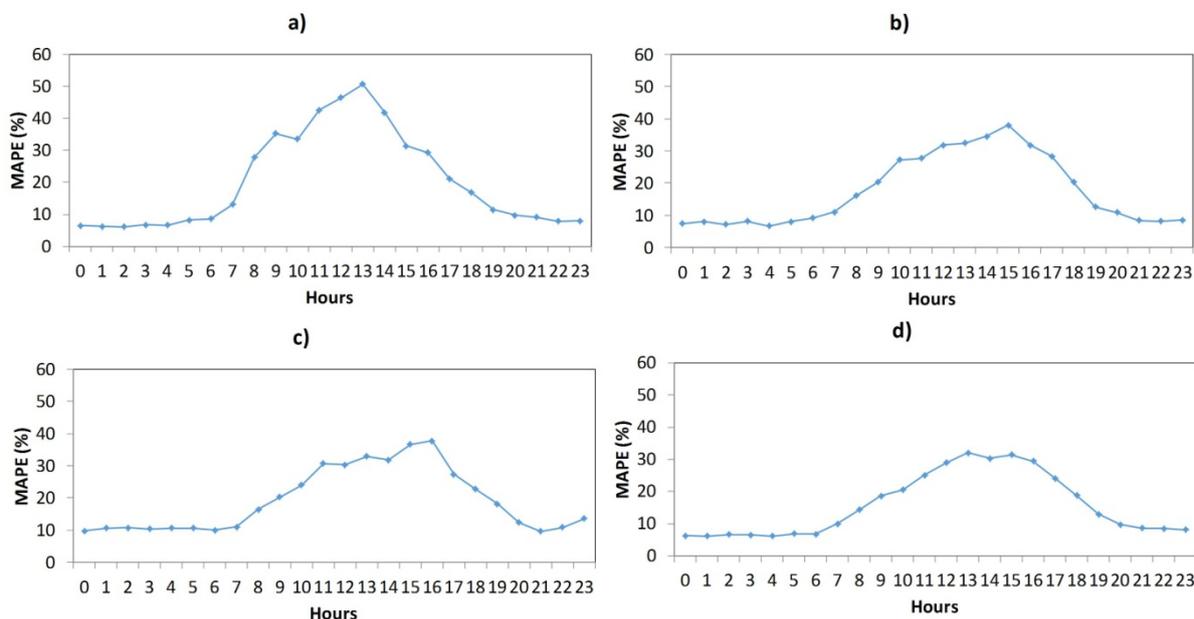


Figure 8 shows the daily average *MAPE* (1) of the winning architecture (all models); the days between 3 August 2012 and 20 August 2012 have a *MAPE* extremely high (over 40%), which is a lot in comparison with the average values. These days, the load profile is an estimate (not actual consumption—the utility makes an estimate), therefore, the Matlab file has employed estimated curves and the *MAPE* is very high. Table 3 shows the days in which consumption was estimated by the utility.

Table 3. Days in which consumption was estimated by the utility.

3 August 2012	Monday
6 August 2012	Friday
10 August 2012	Monday
13 August 2012	Friday
17 August 2012	Monday
20 August 2012	Friday

Most representative forecasted days in which *MAPE* exceeded 40% (days between 3 August 2012 and 20 August 2012) have been represented in Figure 9 and Figure 10. In Figure 9 Mondays are represented, according to the days 6, 13, 20; and Figure 10 shows Fridays, days 3, 10 and 17, respectively. The figures indicate the real consumption of the forecast day (red line, kWh), the day before forecast (green line, kWh), and *MAPE* for those days. The *MAPE* follows a similar trend to the consumption of the previous day which means that if the previous day had a high consumption, the *MAPE* of the forecasting (next day) will be high. In other words, the higher consumption of the day before forecast, the larger the *MAPE* will be.

Figure 8. Daily Average MAPE (%) (1) of the winning architecture: (a) F_L ; (b) F_L_W ; (c) $F_L_W_{DW}$; (d) $F_L_W_M$.

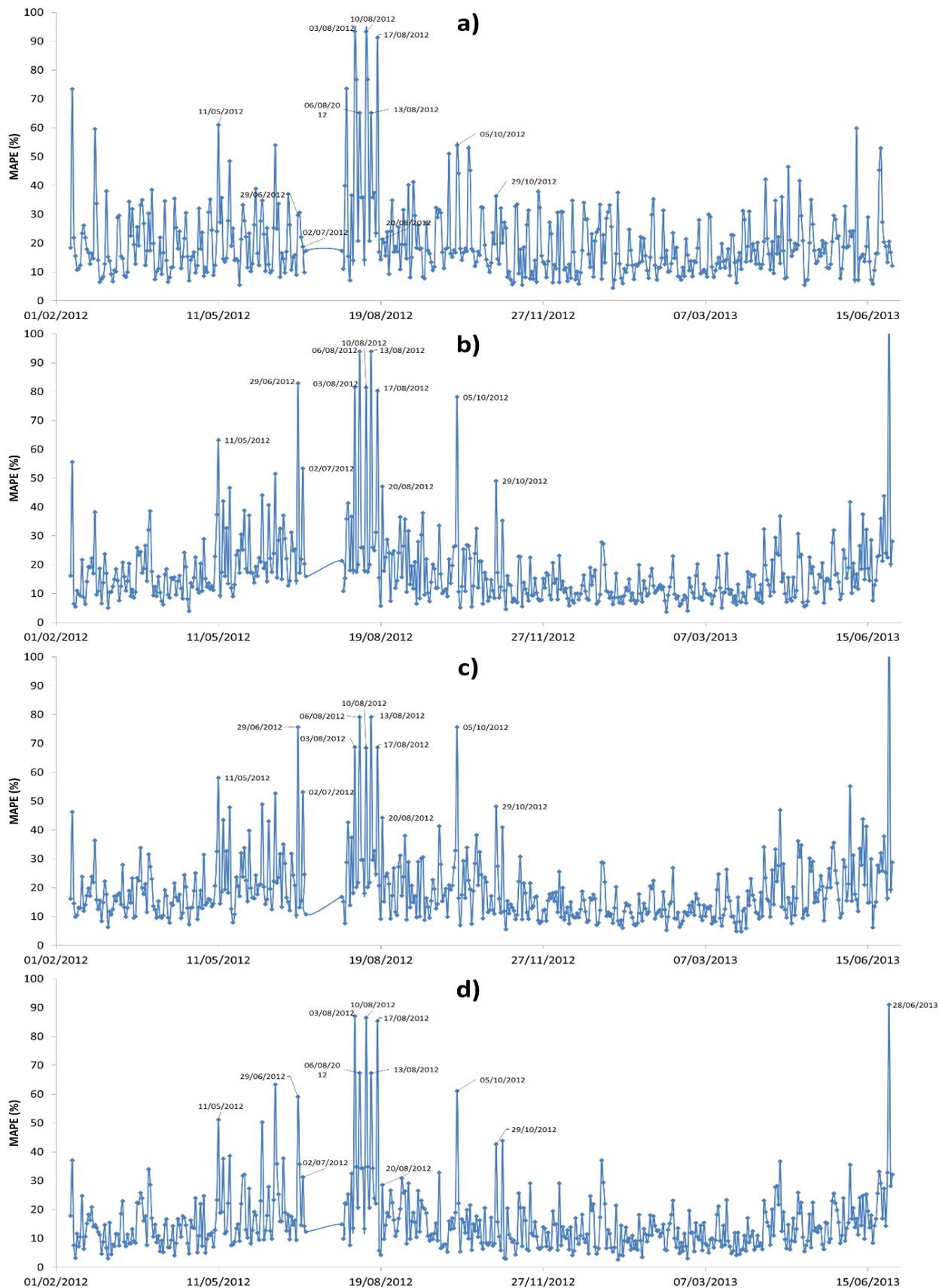


Figure 9. MAPE (%) (1) over 40%, 3–20 August 2012. Mondays: (a) F_L ; (b) F_{L_W} ; (c) $F_{L_W_{DW}}$; (d) $F_{L_W_M}$.

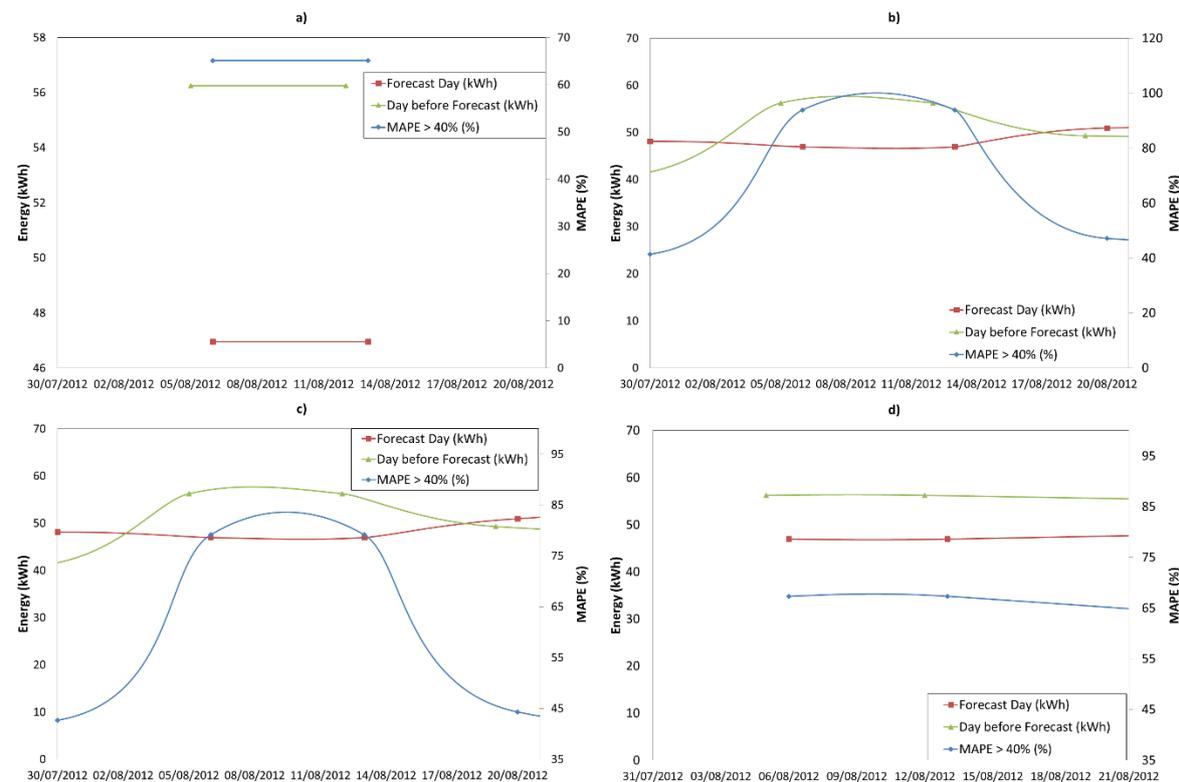


Figure 10. MAPE (%) (1) over 40%, 3–20 August 2012. Fridays: (a) F_L ; (b) F_{L_W} ; (c) $F_{L_W_{DW}}$; (d) $F_{L_W_M}$.

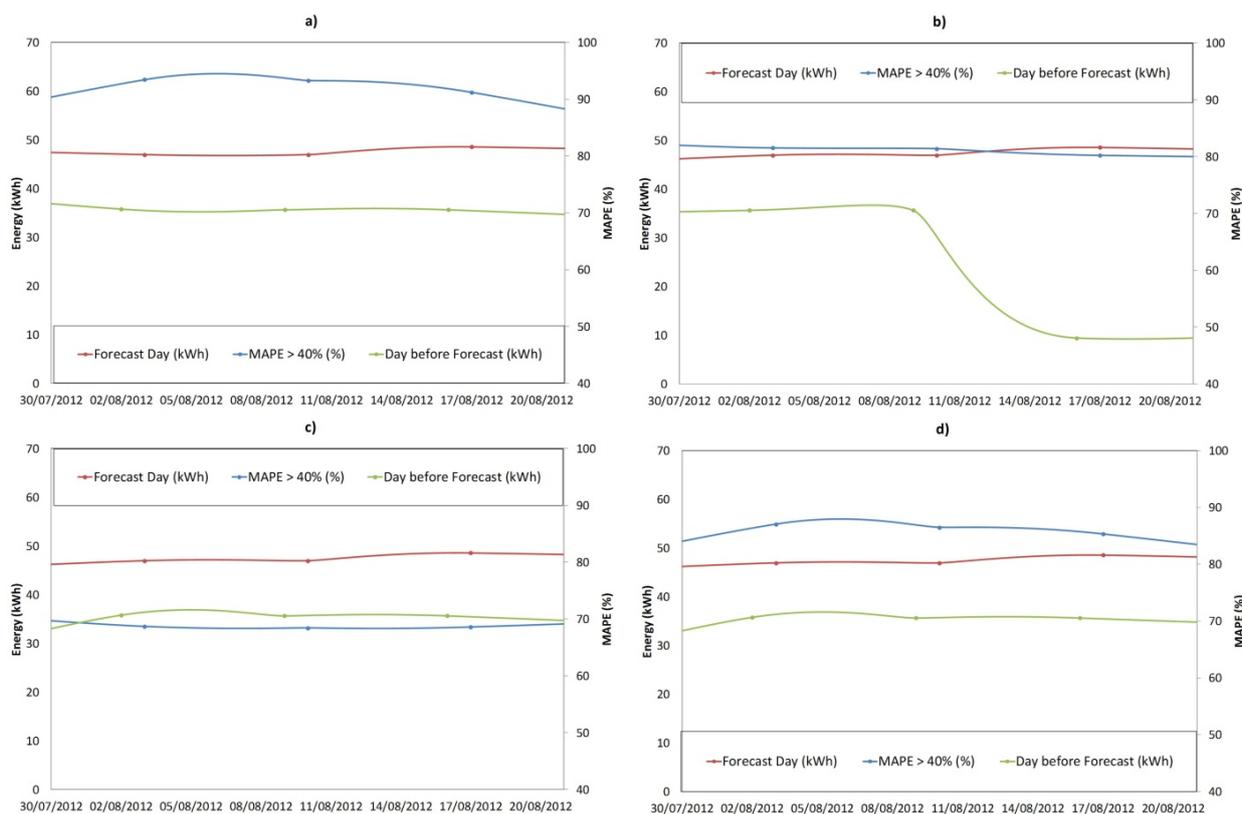
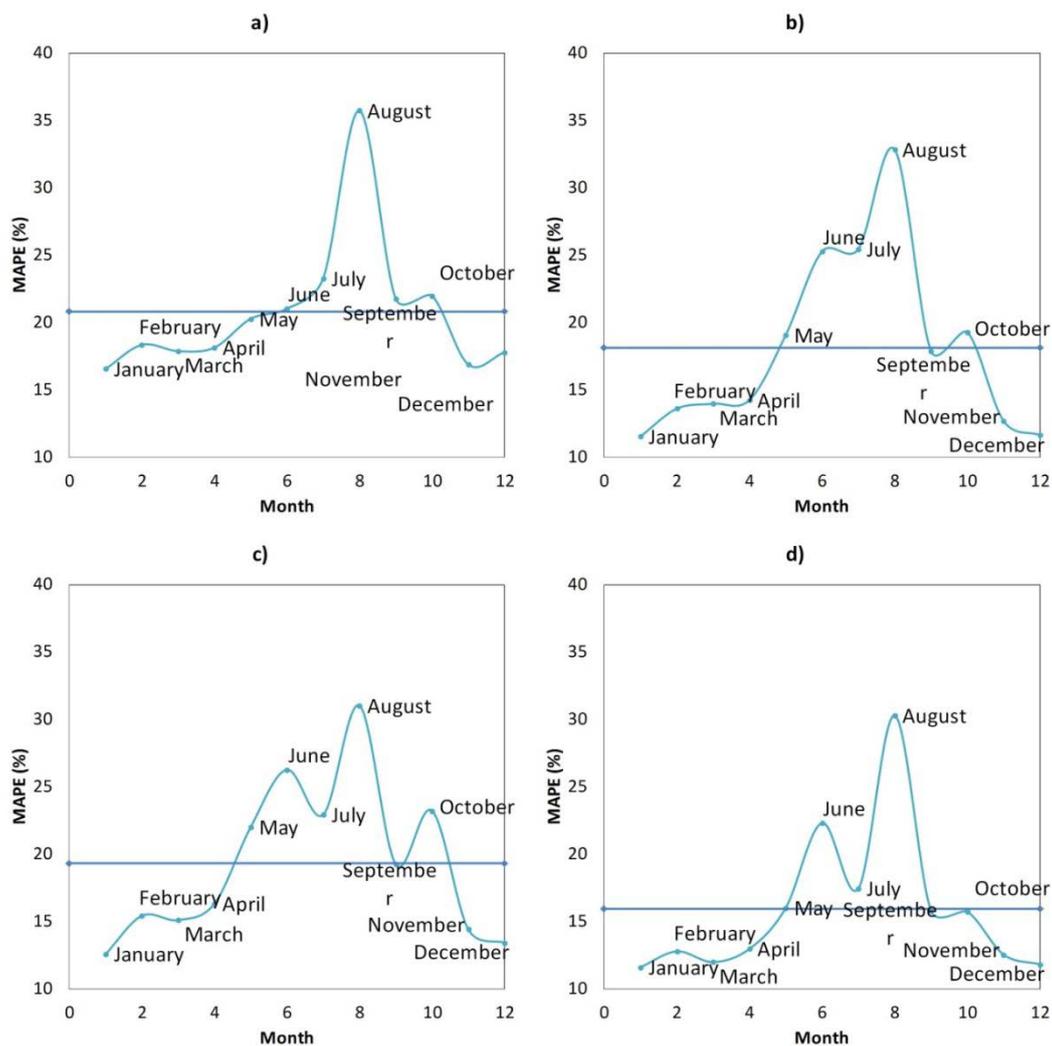


Figure 11 shows the monthly average of the winning architecture (all models). For all of them, the form of the curve ($MAPE$'s evolution) is similar. Months with higher errors are the ones that contain days with load profiles estimated by the utility. The most exact model is $F_L_W_M$, but the most exact model for June is F_L , with a $MAPE$ of 22.20%, whereas the other models have 25.50% (F_L_W), 26.20% ($F_L_W_DW$) and 23.55% ($F_L_W_M$). All the models (except F_L_W) have a low $MAPE$ in winter's months.

Figure 11. Monthly Average $MAPE$ (%) (6) of the winning architecture (the dark blue line represents the Average $MAPE$): (a) F_L ; (b) F_L_W ; (c) $F_L_W_DW$; (d) $F_L_W_M$.



5.2. Result Analysis for Data Set B

Figure 12 shows the hourly average $MAPE$ (3) of the winning architecture (both models). The best one is in line with the results of Table 2, where the lowest $MAPE$ corresponds to $F_L_W_M$ model ($F_L_W_M$ is the best model). In both models, the highest errors occur in the central hours (9–14 h); these hours coincide with the operation of the plants. In case of $F_L_W_M_SR$, the solar radiation makes worsen the forecast (central hours of the day).

Figure 13 shows the daily averaged $MAPE$ (1) of the winning architecture (both models). In both of them, the evolution of the $MAPE$ is similar.

Figure 12. Hourly Average *MAPE* (%) (3) of the winning architecture (all models): (a) *F_L_W_M* (16 neurons, *trainbr*); (b) *F_L_W_M_SR* (26 neurons, *trainbr*).

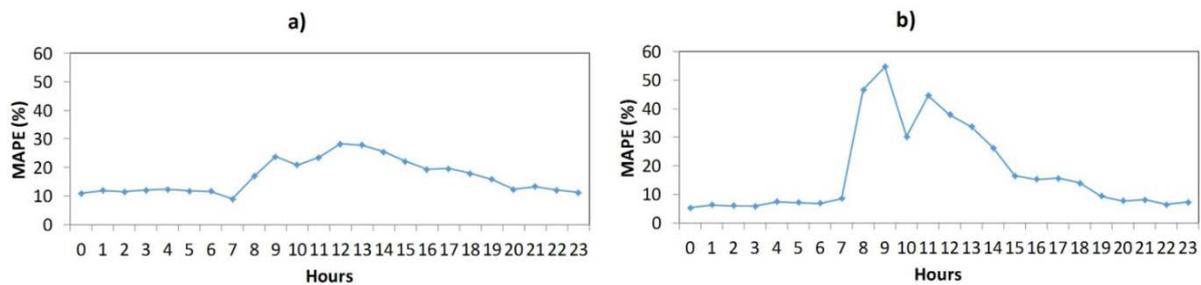
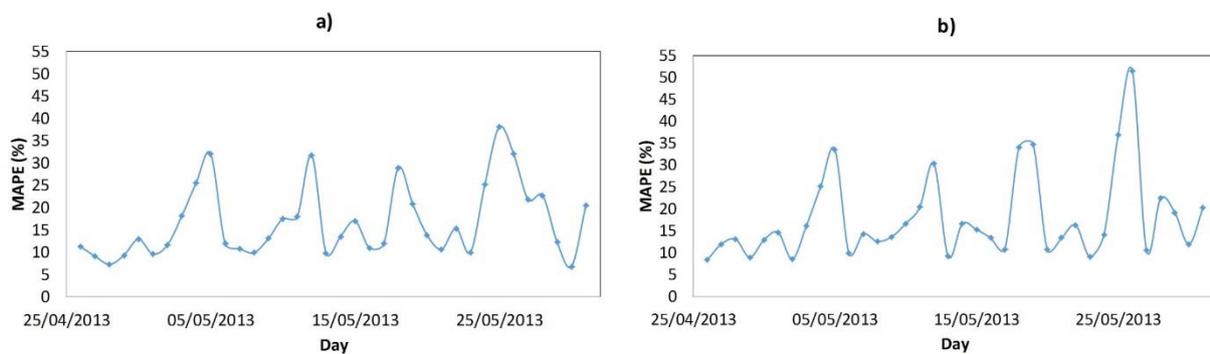


Figure 13. Daily Average *MAPE* (%) (1) of the winning architecture: (a) *F_L_W_M*; (b) *F_L_W_M_SR*.



6. Conclusions and Future Work

Forecasting is more complex in a *microgrid* due to the increased variability of disaggregated load curves. The accuracy of forecasts in a *microgrid* depends on the specific variables employed and the way they are presented to the *ANN*. Several different models, which use different input variables, have been presented along this paper; the most accurate model is *F_L_W_M*, which uses the following variables: load profile, workability and month. Although *CEDER-CIEMAT* has photovoltaic plants, the use of solar radiation as a model input does not improve the prediction

Also, it has been demonstrated that a model trained with a limited number of patterns can obtain error rates similar to another model trained with more patterns. *F_L_W_M* in Data Set A has 15.34% (*MAPE*) and *F_L_W_M* in Data Set B has 16.69% (*MAPE*); Data Set A employs 38 months (approx.) in learning phase and Data Set B employs only 3 months (approx.) in learning phase.

In future works we will test new demand prediction models and generation in disaggregated environments (*microgrids*), in the same line of investigation as Keyhani [25]. Several factors such as predictions using the information of the plants operation will be added. Moreover, we will try to validate a new system in *microgrid*-sized environments (*CEDER-CIEMAT*).

Acknowledgments

Our gratitude to *CEDER-CIEMAT* for providing the data to the presented work. In the same way, we want to convey our gratitude to the project partners *MIREDCON* (IPT-2012-0611-120000), funded by the *INNFACTO* agreement of the Ministry of Economy and Competitiveness of the

Government of Spain. Finally, a special mention to the help of the students Fatih Selim Bayraktar and Guniz Betul Yasar of Gazi University (Turkey), and Cristina Gil Valverde of UNED (Spain).

Author Contributions

Luis Hernandez, Carlos Baladrón, Javier M. Aguiar, Lorena Calavia, Belén Carro, Antonio Sánchez-Esguevillas and Jaime Lloret: definition of the methodology. Luis Hernández, Ángel Fernández and Francisco Pérez: test execution. Luis Hernández, Ángel Fernández and Francisco Pérez: draft. Luis Hernández, Ángel Fernández, Francisco Perez and Jaime Lloret: review.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Spencer, H.H.; Hazen, H.L. Artificial representation of power systems. *Trans. Am. Inst. Electr. Eng.* **1925**, *XLIV*, 72–79.
2. Hamilton, R.F. The summation of load curves. *Trans. Am. Inst. Electr. Eng.* **1944**, *63*, 729–735.
3. Forrest, J.S. The effects of weather on power-system operation. The effects of weather on power system operation. *J. Inst. Electr. Eng. Pt. I Gen.* **1946**, *93*, 161–163.
4. Davies, M. The relationship between weather and electricity demand. *Proc. IEE Pt. C Monogr.* **1959**, *106*, 27–37.
5. Gruetter, J.G. The application of business machines to electrical utility load forecasting. *Trans. Am. Inst. Electr. Eng. Power Appar. Syst. Pt. III* **1955**, *74*, 854–858.
6. Mathewman, P.D.; Nicholson, H. Techniques for load prediction in the electricity-supply industry. *Proc. Inst. Electr. Eng.* **1968**, *115*, 1451–1457.
7. Hippert, H.S.; Pedreira, C.E.; Souza, C.R. Neural networks for short-term load forecasting: A review and Evaluation. *IEEE Trans. Power Syst.* **2001**, *16*, 44–51.
8. Hernández, L.; Baladrón, C.; Aguiar, J.M.; Carro, B.; Sánchez-Esguevillas, A.; Lloret, J.; Massana, J. a survey on electric power demand forecasting: Future trends in smart grids, microgrids and smart buildings. *IEEE Commun. Surv. Tutor.* **2013**, submitted.
9. García-Ascanio, C.; Mate, C. Electric power demand forecasting using interval time series: A comparasion between VAR and iMLP. *Energy Policy* **2009**, *38*, 715–725.
10. Marín, F.J.; Garcia-Lagos, F.; Joya, G.; Sandoval, F. Global model for short-term load forecasting using artificial neural networks. *IEE Proc. Gener. Transm. Distrib.* **2002**, *149*, 121–125.
11. Hernández, L.; Baladrón, C.; Aguiar, J.M.; Carro, B.; Sánchez-Esguevillas, A. Classification and clustering of electricity demand patterns in industrial parks. *Energies* **2012**, *5*, 5215–5228.
12. Hernández, L.; Baladrón, C.; Aguiar, J.M.; Calavia, L.; Carro, B.; Sánchez-Esguevillas, A.; Cook, D.J.; Chinarro, D.; Gómez, J. A Study of the relationship between weather variables and electric power demand inside a smart grid/smart world framework. *Sensors* **2012**, *12*, 11571–11591.

13. Hagan, M.T.; Behr, S.M. The time series approach to short term load forecasting. *IEEE Trans. Power Syst.* **1987**, *2*, 785–791.
14. Badri, M.A. Neural Networks of Combination of Forecasts for Data with Long Memory Pattern. In Proceedings of the IEEE International Conference on Neural Networks, Washington, DC, USA, 3–6 June 1996; Volume 1, pp. 359–364.
15. Hernández, L.; Baladrón, C.; Aguiar, J.M.; Carro, B.; Sánchez-Esguevillas, A.; Lloret, J. Short-term load forecasting for microgrids based on artificial neural networks. *Energies* **2013**, *6*, 1385–1408.
16. Kim, H.; Thottan, M. A two-stage market model for microgrid power transactions via aggregators. *Bell Labs Tech. J.* **2011**, *16*, 101–107.
17. Zhou, L.; Rodrigues, J.J.P.C.; Oliveira, L.M. QoE-driven power scheduling in smart grid: architecture, strategy, and methodology. *IEEE Commun. Mag.* **2012**, *50*, 136–141.
18. Zhou, L.; Rodrigues, J.J.P.C. Service-oriented middleware for smart grid: Principle, infrastructure, and application. *IEEE Commun. Mag.* **2013**, *51*, 84–89.
19. Wille-Haussman, B.; Erge, T.; Wittwer, C. Decentralized optimization of cogeneration in virtual power plants. *Sol. Energy* **2010**, *84*, 604–611.
20. Hernández, L.; Baladrón, C.; Aguiar, J.M.; Carro, B.; Sánchez-Esguevillas, A.; Lloret, J.; Chinarro, D.; Gómez-Sanz, J.J.; Cook, D. A multi-agent system architecture for smart grid management and forecasting of energy demand in virtual power plants. *IEEE Commun. Mag.* **2013**, *51*, 106–113.
21. Drezga, I.; Rahman, S. Phase-Space Short-Term Load Forecasting for Deregulated Electric Power Industry. In Proceedings of International Joint Conference on Neural Networks, Washington, DC, USA, 10–16 July 1999; Volume 5, pp. 3405–3409.
22. Ramezani, M.; Falaghi, H.; Haghifam, M.-R. Short-Term Load Forecasting Using Neural Networks. In Proceedings of the 2005 International Conference on Compute as a Tool (EUROCON 2005), Belgrade, Serbia, 21–24 November; Volume 2, pp. 1525–1528.
23. Razavi, S.; Tolson, B.A. A new formulation for feedforward neural networks. *IEEE Trans. Neural Netw.* **2011**, *22*, 1588–1598.
24. Bishop, C.M. Neural Networks and their applications. *Rev. Sci. Instrum.* **1994**, *65*, 1803–1832.
25. Keyhani, A. *Design of Smart Power Grid Renewable Energy Systems*; John Wiley & Son, Inc.: Hoboken, NJ, USA, July 2011.