# Forecasting Hourly Electricity Load Profile Using Neural Networks

Mashud Rana and Irena Koprinska School of Information Technologies University of Sydney Sydney, Australia {mashud.rana, irena.koprinska}@sydney.edu.au Alicia Troncoso School of Engineering University Pablo de Olavide Seville, Spain atrolor@upo.es

Abstract—We present INN, a new approach for predicting the hourly electricity load profile for the next day from a time series of previous electricity loads. It uses an iterative methodology to make the predictions for the 24-hour forecasting horizon. INN combines an efficient mutual information feature selection method with a neural network forecasting algorithm. We evaluate INN using two years of electricity load data for Australia, Portugal and Spain. The results show that it provides accurate predictions, outperforming three state-of-the-art approaches (weighted nearest neighbor, pattern sequence similarity and iterative linear regression), and a number of baselines. INN is also more accurate and efficient than a noniterative version of the approach. We also found that although the range of load values for the three countries is very different, the load curves show similar patterns, which resulted in more than 90% overlap in the selected lag variables.

Keywords—electricity load prediction; iterative neural network; mutual information

# I. INTRODUCTION

We consider the task of forecasting the hourly electricity load profile for the next day, from a time-series of previous electricity loads. A load profile shows the electricity load (demand) for a given region over a given time period, e.g. a day or month, at a given frequency, typically every hour. Specifically, our task can be stated as follows: given a time series of electricity loads measured every hour up to day *d*, predict the 24 hourly electricity loads for day d+1. This task is categorized as short-term load forecasting.

Forecasting the hourly load profile is needed for the planning and operation of power systems. It is used for making decisions about the dynamic dispatching of generators, setting the minimum reserve to an optimum level and supporting the electricity market participants in their bidding and transactions. The accuracy of the forecasts is very important as overprediction will cause starting too many generators and wasting electricity, while under-prediction may cause disruptions in electricity supply or buying electricity at a high price.

Predicting the hourly load profile with high accuracy is a difficult task. The electricity load time series is complex and non-linear, with superimposed daily, weekly and annual cycles and random components. The random components are due to fluctuations in the electricity usage of individual users, large industrial units with irregular hours of operation, special events and holidays and sudden weather changes.

Various approaches for short-term load forecasting have been proposed. They fall into two groups: statistical approaches, e.g. Autoregressive Integrated Moving Average (ARIMA), exponential smoothing and Linear Regression (LR), and computational intelligence approaches, e.g. Neural Networks (NNs). NNs have received a lot of attention recently due to their ability to learn the time series from examples and to capture the non-linear relationship between the predictor variables and the target variable. Most of the proposed computational intelligence approaches focus on one step ahead forecasting, i.e. at time *t* the task is to predict the load for time t+1. The goal of this paper is to investigate the effectiveness of an iterative NN-based approach for multiple steps ahead forecasting. Our contribution can be summarized as follows:

1) We propose a new approach for forecasting the hourly electricity load profile, called Iterative Neural Network (INN). It combines an efficient implementation of a non-linear feature selector, Mutual Information (MI), with a NN. The predictions are made iteratively, i.e. the prediction for time t+1 is used to make a prediction for time t+2 and so on for all points from the forecasting horizon.

2) We conduct a large-scale evaluation using real electricity data for two years, for three different countries: Australia, Portugal and Spain. We first compare the three time series in terms of range of values, cycles and selected features. We then evaluate the predictive accuracy of INN and compare it with three state-of-the-art load forecasting approaches, WNN [1], PSF [2] and iterative LR, and also with three baselines.

3) We compare the iterative NN approach with a noniterative NN approach. The non-iterative approach uses a separate NN for each hour in the forecasting horizon, while INN uses only one NN. We found that there is no gain in accuracy by using a non-iterative approach.

The rest of the paper is structured as follows. Section II reviews the related work. Section III describes the data and data characteristics. Section IV presents the problem statement and Section V describes the proposed INN approach. Section VI outlines the experimental setup and Section VII presents and discusses the results. Section VIII summarizes the main conclusions.

## II. RELATED WORK

Short-term load forecasting has been an active area of research. There are two main groups of approaches: statistical and computational intelligence. Prominent examples of the first group are exponential smoothing, ARIMA and LR, and notable examples of the second group are NNs and support vector regression.

#### A. Statistical Approaches

The statistical approaches are the traditional approaches for short-term load forecasting. Taylor et al. [3] considered the task of predicting the hourly electricity load for Rio de Janeiro, for lead time from 1 to 24 hours. They compared four methods: ARIMA, double seasonal Holt-Winters exponential smoothing, backpropagation NN and a PCA-based LR. The most accurate method was exponential smoothing which was also the simplest and fastest. To validate these results, in [4] Taylor and McSharry compared the two best methods from their previous work (exponential smoothing and PCA-based LR) with two new methods: an alternative formulation of exponential smoothing and a periodic autoregression, using hourly data for Italy, Norway and Sweden. The results showed that the double seasonal Holt-Winters exponential smoothing was again the best performing method.

Recently, Taylor [5] proposed five novel formulations of exponential smoothing and evaluated them using British and French half-hourly data. The results showed that the best approach was intraday exponential smoothing. Soares and Medeiros [6] proposed a forecasting model consisting of two components: deterministic that describes trends, seasonality and special days, and stochastic that uses linear autocorrelation. A different model was built for each hour of the day. An evaluation using hourly data for Rio de Janeiro showed that the proposed approach obtained good results and outperformed seasonal ARIMA and other benchmark methods.

Fan and Hyndman [7] proposed a semi-parametric additive regression methodology that was used to forecast the halfhourly electricity loads one day ahead for the states of Victoria and South Australia in Australia. A separate model is built for each half hour. It uses previous lagged electricity loads, calendar and temperature variables. The forecasting model showed excellent performance on both historical data and when applied in real time on site.

### B. Computational Intelligence Approaches

NN-based approaches have received considerable attention in load forecasting. NNs are attractive as they can learn the time series from examples and can also model non-linear relationships between the predictor variables and the target variable.

In [8] NNs were used to predict the hourly demands 1-24 hours ahead for North America. A multilevel wavelet transform was used to decompose load into several wavelet components that were then predicted using backpropagation NNs. Chen et al. [9] also considered predicting the demand 1-day ahead from previous hourly data using wavelet and backpropagation NN. They selected a day that is similar to the day to be forecasted in terms of weekly index and weather,

decomposed the demand for it into two wavelet components and then trained a separate backpropagation NN for each component. An evaluation using 4 years data for the region of New England showed the effectiveness of their approach. A similar-day approach is also used in [10]. A weighted Euclidean distance is used to select similar days based on previous load and temperature data. A NN is then trained to predict the correction between the similar day and the day for which the forecast is made. The prediction model was used to forecast Japanese load 1 hour ahead.

There are two prominent approaches that have considered the task of predicting the 24 hourly values for the next day simultaneously: Weighted Nearest Neighbor (WNN) [1, 11] and Pattern Sequence-Based Forecasting (PSF) [2].

Let  $X_i$  be a 24-dimensional vector consisting of the hourly demands for a day *i*. To predict the load for a new day  $X_{d+1}$ , WNN first finds the *k* nearest neighbors of  $X_d$ . The prediction for the new day is the weighted linear combination of the load for the days following the nearest neighbors, where the weights are determined by the distance of the neighbors to  $X_d$ . WNN was applied for predicting both electricity loads and electricity prices and was shown to outperform a number of approaches including NNs and GARCH autoregressive models.

PSF is a generalization of WNN. It combines clustering with sequence matching. It first groups all vectors  $X_i$  from the training data into k clusters and labels them with the cluster number. It then extracts a sequence of consecutive days, from day d backwards, and matches the cluster labels of this sequence against the training data to find a set of sequences that are the same,  $ES_d$ . It then follows a nearest neighbor approach similar to WNN - finds the following day for each element of  $ES_d$  and averages the 24 hourly loads of these following days, to produce the final 24 hourly predictions for day d+1. The results showed that PSF is a very competitive approach outperforming ARIMA, support vector regression and NNs.

#### III. DATA AND DATA CHARACTERISTICS

We use electricity load data for three different countries: Australia, Portugal and Spain, for two years: 2010 and 2011 (from 1 January 2010 to 31 December 2011). The data is sampled every hour, and the total number of samples in each dataset is 365\*24\*2=17,520.

All data is publicly available. The Australian data is for the state of NSW; it is provided by the Australian Energy Market Operator (AEMO) and available at [12]. The Portuguese and Spanish data are provided by the Spanish Electricity Price Market Operator (OMEL) and are available at [13].

The electricity load has three main cycles: daily, weekly and yearly. Fig. 1 shows the hourly load for the three countries for the same fortnightly period - from Monday 14 June to Sunday 27 June, 2010. The graphs are very similar and clearly show the daily and weekly cycles. The daily cycle is 24 hours and the weekly cycle is 7\*24=168 hours.

The daily pattern is evident from the similarity of the load profiles of the individual days, e.g. the profile of Monday is similar to the profile of Tuesday. The load is lowest at around 4:30am; it then increases and reaches its first peak at 9-9.30am, stays relatively stable till the end of the working day and then reaches its second peak at 6-7pm (showing more irregular ups and downs for the Portuguese and Spanish data), before gradually decreasing, which is consistent with the industrial, business and human daily routine. The weekly pattern is evident from the similarity of load profiles of the two weeks – we can see that the load during the weekdays (Monday to Friday) is higher than the load during the weekends (Saturday and Sunday), which is as expected. We detect and use the daily and weekly patterns to conduct feature selection.

There is also a yearly pattern of the electricity load – e.g. the load for 2010 is very similar to the load for 2011. This motivates using the 2010 data to build prediction models and then using these models to predict the load for 2011.



Fig. 1. Hourly electricity load for two consecutive weeks for Australia, Portugal and Spain

Although the load curves for the three countries are similar due to the similar cyclic nature of the electricity load, it is important to note that the range of load values differs considerably. For the 2-week period shown on Fig.1, we can see that the range of load values for Australia is from 7,000 to 12,000 MW, for Portugal – from 2,000 to 6,000 MW and for Spain – from 14,000 to 27,000 MW. Thus, the load for Spain is the highest, followed by Australia and Portugal.

#### IV. PROBLEM STATEMENT

Given a time series of hourly electricity loads up to day d, the goal is to forecast the 24 hourly loads for day d+1, or the hourly load profile for day d+1. More formally, given a time sequence of days  $X_1, X_2, ..., X_d$ , where each element is a vector of hourly electricity loads, e.g.  $X_i = (X_{i,1},..,X_{i,24})$  is the hourly load profile for day i, the goal is to forecast  $X_{d+1} = (X_{d+1,1},..,X_{d+1,24})$ , the hourly load profile for day d+1.

Therefore, the forecasting is based only on previous electricity loads and it ignores the weather conditions. This is consistent with previous studies - weather conditions such as temperature and humidity are rarely used for prediction horizons of one day ahead or less. The main reason for this is that the changes in the weather variables for such small horizons are already captured in the load series. However, recent studies using British electricity data [5, 14] found that although the previous load data was sufficient for making forecasts up to 5 hours ahead, the use of weather variables as additional features was beneficial for longer time horizons. It is also possible that the weather variables are more useful for some countries than others, e.g. for countries with more sudden weather changes. In future work we plan to add weather variables and investigate if they are useful for load prediction for Australia, Portugal and Spain.

# V. PROPOSED APPROACH - ITERATIVE NEURAL NETWORK

The key idea of the proposed approach is to build an iterative prediction model. At time *t* it will make a prediction for time t+1; this prediction will be added to the training data, and used to make a prediction for time t+2 and so on for all points from the forecasting horizon. We chose to combine a nonlinear prediction model, NN, with a non-linear feature selector, MI. The proposed approach is called Iterative Neural Network (INN). Below we describe its main steps: feature selection and building of the prediction model.

#### A. Feature Selection

Appropriate feature selection is one of the crucial factors for successful modeling and forecasting of time series data. Most of the existing approaches for short-term load forecasting use correlation-based techniques and heuristics that only capture the linear relationship between the target variable and the predictor variables [3-5, 14-16]. In this paper we investigate the use of a feature selector that captures both linear and non-linear dependencies - MI.

MI measures the interdependence between two random variables. If the two variables are independent, MI is zero; if they are dependent, MI has a positive value reflecting the strength of their dependency. We apply a novel and fast method for computing MI, based on *k*-nearest neighbor distances [17]. It was shown to be efficient and more reliable than the traditional MI methods.

The MI between two random continuous variables *X* and *Y* with dimensionality *N* is estimated as:

$$MI(X,Y) = \psi(k) - \frac{1}{k} - \frac{1}{n} \sum_{i=1}^{N} (\psi(n_x(i)) + \psi(n_y(i))) + \psi(N)$$

where  $\psi(x)$  is the digamma function, k is the number of nearest neighbors (we used k = 6);  $n_x(i)$  is the number of points  $x_j$  with a distance to  $x_i$  satisfying  $||x_i - x_j|| \le \varepsilon_x(i)/2$  and  $n_y(i)$  is the number of points  $y_j$  with a distance to  $y_i$  satisfying  $||y_i - y_j|| \le \varepsilon_y(i)/2$ , where  $\varepsilon_x(i)/2$  is the distance between  $x_i$  and its k-th neighbor in the X subspace and  $\varepsilon_y(i)/2$  is the distance between  $y_i$  and its k-th neighbor in the Y subspace.

To conduct feature selection using MI, we first form a set of candidate features that includes all lag variables from a 1week sliding window. We found that 1 week is a sufficient length to capture the strong daily and weekly patterns in the load data. Since we are using hourly data, the candidate set consists of 7\*24 = 168 features. We then compute the MI score between each of the 168 candidate features and the target variable, and rank the candidate features in decreasing order based on their MI score. Fig. 2 plots the normalized MI score for each feature in ranked order, for the three data sets.

Fig. 2 shows that the three graphs are very similar. The MI curve shows a sharp drop at the beginning, followed by a more gradual decrease after rank 10 and no significant changes after rank 50. Based on these result, we select the 50 highly ranked features for each dataset. These features form the final feature sets for each country. They are denoted as  $FS_{Australia}$ ,  $FS_{Portugal}$  and  $FS_{Spain}$  and are listed in Table I.



Fig. 2. MI score and feature ranking

Table I shows that there is a considerable overlap between the three feature sets – 90% between  $FS_{Australia}$  and  $FS_{Portugal}$ , 94% between  $FS_{Australia}$  and  $FS_{Spain}$  and 94% between  $FS_{Portugal}$ and  $FS_{Spain}$ . The selected features also reflect the daily and weekly cycle of the electricity load – lag 24 and multiples of 24 (daily), and also lag 168 (weekly). Hence, the feature selection confirms the similar cyclic nature of the three time series, despite the different ranges of their values.

TABLE I. FEATURE SETS

Feature set	Selected lag variables to predict the load at hour $h, X_h$
FS <sub>Austalia</sub>	$X_{h-1}$ to $X_{h-6}$ ; $X_{h-20}$ to $X_{h-28}$ ; $X_{h-45}$ to $X_{h-51}$ ; $X_{h-70}$ to $X_{h-75}$ ;
	$X_{h-94}$ to $X_{h-99}$ ; $X_{h-118}$ to $X_{h-122}$ ; $X_{h-141}$ to $X_{h-147}$ ; $X_{h-165}$ to
	X <sub>h-168</sub>
FS <sub>Portugal</sub>	$X_{h-1}$ to $X_{h-9}$ ; $X_{h-20}$ to $X_{h-28}$ ; $X_{h-46}$ to $X_{h-50}$ ; $X_{h-70}$ to $X_{h-74}$ ;
_	$X_{h-82}$ ; $X_{h-86}$ ; $X_{h-95}$ to $X_{h-98}$ ; $X_{h-118}$ to $X_{h-122}$ ; $X_{h-141}$ to
	$X_{h-147}$ ; $X_{h-165}$ to $X_{h-168}$
FS <sub>Spain</sub>	$X_{h-1}$ to $X_{h-9}$ ; $X_{h-20}$ to $X_{h-27}$ ; $X_{h-46}$ to $X_{h-51}$ ; $X_{h-70}$ to $X_{h-74}$ ;
	$X_{h-94}$ to $X_{h-98}$ ; $X_{h-118}$ to $X_{h-122}$ ; $X_{h-141}$ to $X_{h-147}$ ; $X_{h-164}$ to
	$X_{h_{2},168}$

#### **B.** Prediction Model

As a prediction model we use a multi-layer perceptron NN, trained with the Levenberg-Marquardt algorithm. We build and train three NNs; one for each data set. Each NN has 50 input neurons corresponding to the features from Table I, one hidden layer (where the number of neurons is set using validation set as described below) and one output neuron. The output neuron corresponds to the load value one step ahead.

## 1) Iterative Prediction

The NN is trained to predict one step ahead - at hour h, it predicts the load for hour h+1. To predict the load for the 24 hours in the forecasting window, it is used iteratively by considering the predicted load in the previous steps as actual load. That is, the predicted value for h+1 is considered as actual value, appended at the end of the available data and used to predict the value for h+2. This continues until the predictions for the desired forecasting horizon are made. More generally, to make a prediction for h+1 of day d+1, where h>1, we use the previously recorded actual data for all days till day d and the previously forecasted data for the hours from 1to h for day d+1.

We note that the feature selection is conducted once only. This is because INN is trained for one step ahead prediction, although it is used iteratively for more than one step ahead prediction.

# 2) Number of Hidden Neurons

The number of neurons in the hidden layer is selected using a validation set. The given data is split into three nonoverlapping datasets: training, validation and testing, see section VI-A. We create P different NNs by varying the number of neurons in the hidden layer from I to P, train each NNs using the training set and evaluate its performance on the validation set. The best NN is then selected; it is the one that provides the most accurate predictions on the validation set. This best NN is then evaluated on the testing data.

## 3) Training Algorithm and Stopping Criterion

To train INN we use the Levenberg-Marquardt algorithm [18]. We chose it over the standard steepest gradient descent backpropagation algorithm due to its faster convergence. The Levenberg-Marquardt algorithm combines the steepest gradient descent and the Gauss-Newton algorithm, switching between

them based on the complexity of the error surface. It preserves the advantages of both algorithms: the fast convergence of the Gauss-Newton algorithm and the stability of the steepest gradient descent when used with a small learning rate. The maximum number of training epochs was set to 2000. The training stopped if there was no improvement in the error for 20 consecutive epochs. We used tangent sigmoid transfer functions for the hidden neurons and a linear transfer function for the output neuron.

## VI. EXPERIMENTAL SETUP

#### A. Data Sets

The available data for each country is divided into three non-overlapped subsets - *training set* ( $D_{train}$ ), *validation set* ( $D_{valid}$ ) and testing set ( $D_{test}$ ).  $D_{train}$  is used for feature selection and building of the prediction models and contains 70% of the data for 2010.  $D_{valid}$  is used for selecting the best NN architecture and contains 30% of the data for 2010.  $D_{test}$  is used for evaluating the predictive accuracy of the models and contains the data for 2011.

#### B. Measuring Predictive Accuracy

To measure predictive accuracy, we use two standard performance measures: Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE), defined as:

$$MAE = \frac{1}{N} \frac{1}{H} \sum_{i=1}^{N} \sum_{h=1}^{H} \left| X_{i,h} - \hat{X}_{i,h} \right|$$
$$MAPE = \frac{1}{N} \frac{1}{H} \sum_{i=1}^{N} \sum_{h=1}^{H} \left| \frac{X_{i,h} - \hat{X}_{i,h}}{X_{i,h}} \right|. 100\%$$

where  $X_{i,h}$  and  $\hat{X}_{i,h}$  are the actual and predicted loads for day *i* at hour *h*, respectively, *N* is the number of samples (days) in the test data and *H* is 24, the number of predicted loads, one for each hour.

MAPE is an extension of MAE. It can be seen as a normalized version of MAE where the normalization is achieved by a simple division by  $X_{i,h}$ . It is more appropriate for comparison of time series with different ranges, as in our case.

### C. Forecasting Methods Used for Comparison

In addition to measuring the predictive accuracy of our approach, it is also important to compare its performance with other forecasting methods. We use three state-of-the-art forecasting methods and also three baselines (naïve forecasting methods) for comparison.

### 1) State-of-the-art Methods

We use the following prominent methods for short-term load forecasting: PSF [2], WNN [1] and LR.

PSF and WNN are two recently proposed methods, applied to the same task as ours. In contrast to our iterative approach, these methods predict the 24 values for the next day simultaneously. PSF uses similarity between sequences of cluster labels and WNN uses weighted nearest neighbors. Both approaches have been shown to be very successful, outperforming a number of statistical and NN methods, on the tasks of electricity load and electricity price prediction.

We also compare our INN approach with a standard statistical method, LR, which we apply iteratively, in the same way as INN; we call this method Iterative LR (ILR). ILR uses a stepwise regression for feature selection. It starts with all features and at each step tests the removal of each feature using the M5 criterion. The feature that improves the model the least is deleted and the process continues until there is no further improvement. We chose this stepwise regression version as it was very successful in our previous study [16]. However, in [16] the task was different - one step ahead prediction of 5-minute load data, and LR was applied in a standard, non-iterative way, while here we use it iteratively.

## 2) Baselines

We implemented the following baselines:

- B<sub>pday</sub>: load from the previous day at the same time. The prediction for X<sub>i,h</sub> is given by X<sub>i-1,h</sub>.
- $B_{pweek}$ : load from the previous week at the same time. The prediction for  $X_{i,h}$  is given by  $X_{i-7,h}$ .
- B<sub>mean</sub>: mean load value in the training data for hour h. The prediction for X<sub>i,h</sub> will be mean(X<sub>j,h</sub>) over all days j in the training data.

#### VII. RESULTS AND DISCUSSION

#### A. Overall Performance

Table II shows the accuracy results, MAE and MAPE, of the proposed INN approach on the three data sets for each hour of the forecasting horizon.

TABLE II. PREDICTIVE AC	CCURACY OF	INN
-------------------------	------------	-----

	Australian data		Portugu	ese data	Spanish data		
Hour	MAE	MAPE	MAE	MAPE	MAE	MAPE	
1	51.12	0.65	163.47	4.66	717.26	3.59	
2	69.50	0.93	200.05	6.31	838.78	4.71	
3	74.10	1.06	230.10	7.76	904.57	5.51	
4	78.88	1.18	250.26	8.79	973.27	6.21	
5	99.49	1.49	261.80	9.37	1086.94	7.11	
6	155.60	2.19	264.70	9.63	1126.89	7.40	
7	237.28	3.01	267.85	9.89	1057.32	6.73	
8	286.25	3.32	295.11	10.48	1201.62	7.08	
9	290.51	3.15	330.21	11.03	1400.63	7.77	
10	298.78	3.14	441.05	12.77	1376.26	7.07	
11	318.89	3.33	503.64	13.37	1240.25	5.92	
12	360.18	3.77	501.71	12.59	1199.07	5.56	
13	405.25	4.26	522.88	12.56	1239.62	5.51	
14	449.27	4.74	523.14	12.83	1230.08	5.46	
15	479.30	5.10	541.03	13.59	1203.82	5.44	
16	495.98	5.25	576.37	14.48	1227.16	5.67	
17	506.70	5.30	584.04	14.96	1256.19	5.94	
18	487.62	5.02	582.84	15.35	1254.80	5.89	
19	443.01	4.53	558.04	14.70	1296.33	5.95	
20	424.34	4.40	535.68	13.54	1282.55	5.85	
21	401.44	4.30	525.29	12.91	1168.14	5.18	
22	358.95	4.01	527.30	12.96	1085.85	4.74	
23	297.48	3.45	526.81	13.03	981.09	4.35	
24	247.54	3.02	520.92	13.29	882.68	4.24	
Avg.	304.89	3.36	426.43	11.70	1134.63	5.79	

Table III compares the INN results with the three state-ofthe-art approaches and the three baselines. Table IV shows the pair-wise comparison of all prediction methods and baselines for statistical significance of differences in accuracy using the Wilcoxon rank sum test.

TABLE III. PREDICTIVE ACCURACY - COMPARISON WITH STATE-OF-THE-ART METHODS AND BASELINES

Prediction	Austral	lian data	Portugu	iese data	Spanish data	
method	MAE	MAPE	MAE	MAPE	MAE	MAPE
INN	304.89	3.36	426.43	11.70	1134.63	5.79
WNN	307.46	3.40	538.87	14.95	1179.89	6.03
PSF	352.03	3.96	589.77	16.18	1711.38	8.87
ILR	332.71	3.69	427.00	11.78	1306.62	6.59
B <sub>pday</sub>	420.46	4.82	579.61	16.06	1888.02	9.47
B <sub>pweek</sub>	471.20	5.20	695.38	19.12	1460.07	7.45
B <sub>mean</sub>	719.44	8.08	653.55	18.11	2671.69	14.32

The main results can be summarized as follows:

- The proposed INN is the most accurate forecasting method for all datasets (Table III). It outperforms all state-of-the-art methods used for comparison on all datasets, and these differences are statistically significant in all but one case (Table IV), see below. The second best approach is WNN, followed by ILR and PSF.
- INN also considerably outperformed all three baselines, on all datasets (Table III). Comparing the baselines, B<sub>pday</sub> is the most accurate baseline overall, followed by B<sub>pweek</sub> and B<sub>mean</sub>.
- All pair-wise differences in accuracy are statistically significant except the difference between INN and ILR on the Portuguese data (Table IV). The p-values are: p≤0.05 for 3 cases and p≤0.001 for all other cases.
- There are differences between the predictive accuracies for the three countries (Table III). The Australian data is predicted most accurately, e.g. INN achieves MAPE<sub>Australia</sub>=3.36%. To better understand what a MAPE with this magnitude means, Fig. 3 shows the actual and predicted load by INN for a typical fortnight period; we can see that the two load curves are considerably overlapping. On the other hand, the Portuguese and Spanish data are predicted with lower accuracy, e.g. INN achieves MAPE<sub>Portugal</sub>=11.70% and MAPE<sub>Spain</sub>=5.79%.
- The Australian data has the lowest MAE and MAPE (Table III). The Spanish data has the highest MAE (MAE<sub>Spain</sub>=1134.63) but the Portuguese data has the highest MAPE (MAPE<sub>Portugal</sub>=11.70). This lack of correlation between the MAE and MAPE rankings for Spain and Portugal is interesting. It is due to the different magnitudes of the hourly electricity load the mean value of the Spanish load is much higher (20810 MW) compared to the mean value of the Portuguese load (1535 MW). Hence, the denominator of the MAPE equation is much higher for Spain than for Portugal, and as a result MAPE<sub>Spain</sub> < MAPE<sub>Portugal</sub> although MAE<sub>Spain</sub> > MAE<sub>Portugal</sub>. As mentioned before, MAPE provides a

normalized error estimate and is more appropriate when comparing results for data with different ranges. Based on this, we can conclude that the accuracy for the Spanish data is higher than the accuracy for the Portuguese data.

TABLE IV. PAIR-WISE STATISTICAL SIGNIFICANCE COMPARISON FOR MAE
AND MAPE (WILCOXON RANK SUM TEST); ** - STAT. SIGN. AT $P \le 0.001$ ,
* - Stat. Sign. at p≤0.05. x - No Stat. Sign. Difference

AUSTRALIAN DATA

	INN	WNN	PSF	ILR	B <sub>pday</sub>	B <sub>pweek</sub>	Bmean
INN		**	**	**	**	**	**
WNN			**	**	**	**	**
PSF				**	**	**	**
ILR					**	**	**
B <sub>pday</sub>						**	**
Bpweek							**
Bmean							

PORTUGUESE DATA

	INN	WNN	PSF	ILR	Bpday	Bpweek	Bmean
INN		**	**	х	**	**	**
WNN			**	**	**	**	**
PSF				**	х	**	**
ILR					**	**	**
Bnday						**	**
B <sub>pweek</sub>							*
Bmaan							

SPANISH DATA

	INN	WNN	PSF	ILR	B <sub>pday</sub>	Bpweek	Bmean
INN		*	**	**	**	**	**
WNN			**	**	**	**	**
PSF				**	*	**	**
ILR					**	**	**
B <sub>pday</sub>						**	**
Bpweek							**
Bmean							



Fig. 3. Actual load and predicted load by INN for the Australian data for a fortnight period (Monday 4 April – Sunday 17 April 2011)

### B. Hourly Performance

Table II also shows the predictive accuracy (MAE and MAPE) for each of the 24 hours in the forecasting horizon and Fig. 4 shows a graphical representation of the MAPE results.



Fig. 4. Hourly MAPE for INN

We can see that the predictive error is different for the different hours of the day. For the Australian data, the error is low at the beginning of the day, it then gradually increases and reaches a peak at about 8.30am and then a second (highest) peak at about 5pm before gradually decreasing. The graph for the Portuguese data is similar but there are two peaks in predictive error – at 11am and 6pm, and then a decrease from 9pm. For the Spanish data, the error is also low at the beginning of the day, it reaches a peak at about 9am, flattens till about 8pm and then decreases. Overall, the predictive error is higher during peak hours, when the electricity demand is higher and more irregular. It is lower when the demand is lower (during the night – from 10pm to 6am), which is consistent with the industrial activity and human routine.

#### C. Iterative vs Non-Iterative approach

To evaluate the effectiveness of the proposed iterative methodology, we compare the performance of the iterative NN approach with a non-iterative NN approach. The non-iterative approach is implemented as follows: 24 different NNs are created; each of them is trained to predict the load for one of the 24 hours in the forecasting horizon, i.e. the first NN predicts the load for h = 1, the second NN for h = 2 and so on. For training and parameter tuning of the 24 NNs we used the same training algorithm, stopping condition and validation set procedure, as used in the iterative approach and described in Section V.

Table V presents the predictive accuracy results of the noniterative NN approach. Fig. 5 shows a comparison of the iterative and non-iterative approaches in terms of the average MAPE. We can see that the iterative approach outperforms the non-iterative approach for all datasets. The average MAPEs are: 4.44% (non-iterative) and 3.36% (iterative) for the Australian data, 18.85% (non-iterative) and 11.70% (iterative) for the Portuguese data and 9.73% (non-iterative) and 5.79% (iterative) for the Spanish data. Thus, the improvement in MAPE of the iterative approach is 24.32%, 37.93% and 40.49% for the three datasets, respectively. The improvement in MAE is 23.92% for the Australian data and 39.49% for the Portuguese and Spanish data.

TABLE V. PREDICTIVE ACCURACY OF THE NON-ITERATIVE NN APPROACH

	Australian data		Portugu	ese data	Spanish data	
Hour	MAE	MAPE	MAE	MAPE	MAE	MAPE
1	98.04	1.24	320.51	9.06	1510.28	7.62
2	101.19	1.37	428.95	13.13	1666.09	9.81
3	101.28	1.46	510.11	16.52	2113.02	13.65
4	123.62	1.83	399.87	14.13	1791.10	11.86
5	141.25	2.10	486.49	17.21	2062.46	14.21
6	216.93	3.05	491.17	17.57	1844.05	12.69
7	339.75	4.32	517.09	19.07	1954.15	12.74
8	402.84	4.66	480.87	16.90	1822.33	10.44
9	386.86	4.19	559.08	18.81	2063.04	11.19
10	405.37	4.31	642.37	18.28	1983.04	9.90
11	403.91	4.31	669.59	17.27	1905.99	9.12
12	535.12	5.72	729.84	17.66	1921.38	8.79
13	586.83	6.33	776.72	18.40	2366.10	10.48
14	570.48	6.17	906.53	20.99	1758.97	7.70
15	540.06	5.69	843.91	20.48	1701.17	7.63
16	689.53	7.39	834.84	19.93	1889.93	8.77
17	697.52	7.33	1023.3	25.42	2032.80	9.34
18	645.28	6.60	985.81	24.71	2172.73	9.96
19	546.74	5.52	966.29	24.67	1712.89	7.87
20	505.42	5.22	858.54	21.11	1934.79	8.87
21	508.40	5.45	870.12	20.16	1895.54	8.66
22	436.78	4.86	866.02	19.62	1846.03	8.06
23	345.52	3.99	864.35	19.60	1521.09	6.84
24	289.42	3.48	881.36	21.81	1532.49	7.42
Avg.	400.76	4.44	704.74	18.85	1875.06	9.73

A comparison of the hourly predictive accuracy of the iterative and non-iterative prediction approaches (see Table II and Table V) shows that the iterative approach is more accurate for all 24 hours of the foresting horizon, on all three datasets. The MAPE of INN for the 24 hours is in the range of 0.65-5.30% for the Australian data, 4.66-15.35% for the Portuguese data and 3.59%-7.77% for the Spanish data, where these values for the non-iterative approach are: 1.24-7.39%, 9.06-25.42% and 6.84-14.21%.

Besides higher accuracy, another important advantage of the iterative approach is that it is simpler and faster to train. INN creates and trains only one NN, while the non-iterative approach uses 24 NNs. The training times were 80 seconds for the single NN of the iterative approach and 32 minutes for the 24 NNs of the non-iterative approach.



Fig. 5. Comparison of the iterative and non-iterative NN approaches - MAPE

To sum up, our results show that the iterative approach, INN, outperforms the non-iterative version in terms of predictive accuracy. No gain in accuracy is obtained by using a separate prediction model for each of the hours in the forecasting horizon. The iterative approach is also faster to train as it uses one NN, compared to 24 NNs for the non-iterative approach.

## VIII. CONCLUSION

We considered the task of predicting the hourly electricity load profile for the next day, from a time series of previous hourly electricity loads. We developed INN, an iterative approach that combines an efficient MI feature selection algorithm with a NN.

We conducted an extensive performance evaluation using electricity data for two years from three different countries: Australia, Portugal and Spain. We found that although the range of load values for the three countries varied significantly, the load curves showed similar cycles. The MI feature selection algorithm was able to successfully extract features that capture the daily and weekly cycles of the electricity load and there was more than 90% overlap between the selected features for the three countries.

We compared the performance of our approach with three state-of-the-art approaches, WNN, PSF and ILR, and three baselines. The results showed that INN was the most accurate approach overall. It outperformed all baselines, and achieved statistically significant improvement over all state-of-the-art methods, for all datasets, except ILR on the Portuguese data where the improvement was not statistically significant. INN achieved average MAPE of 3.36% for the Australian data, 11.70% for the Portuguese data and 5.79% for the Spanish data. The predictive error was higher for the peak hours and lower for the night hours.

We also conducted a comparison of the proposed iterative NN approach with a non-iterative NN approach, where a separate NN is used to predict the load for each hour of the forecasting horizon. The iterative approach was not only more computationally efficient (as it uses only one NN) but it also outperformed the non-iterative approach in terms of predictive accuracy – the improvement was from 24.32% to 40.49% in MAPE.

Therefore, we conclude that the proposed INN approach is a viable option for forecasting the hourly electricity load profile – it provides high accuracy and has low computational requirements. The WNN and ILR approaches also showed good results, and WNN has an advantage as it predicts the 24 values for the next day simultaneously.

In future work we plan to investigate the effectiveness of INN for forecasting of other energy time series, e.g. solar and wind power and electricity prices. We will also study the performance of INN for forecasting horizons longer than 24 hours.

#### ACKNOWLEDGMENT

Mashud Rana is supported by an Endeavour Postgraduate Award funded by the Australian Government. This work was also supported in part by the Junta de Andalucía under project P12-TIC-1728 and by the University Pablo de Olavide under project APPB813097.

#### References

- A. Troncoso, J. M. Riquelme, J. C. Riquelme, J. L. Martinez, and A. Gomez, "Electricity market price forecasting based on weighted nearest neighbor techniques," *IEEE Transactions on Power Systems*, vol. 22, pp. 1294-1301, 2007.
- [2] F. Martínez-Álvarez, A. Troncoso, J. C. Riquelme, and J. S. Aguilar-Ruiz, "Energy time series forecasting based on pattern sequence similarity," *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, pp. 1230-1243, 2011.
- [3] J. W. Taylor, L. M. d. Menezes, and P. E. McSharry, "A comparison of univariate methods for forecasting electricity demand up to a day ahead," *International Journal of Forecasting*, vol. 22, pp. 1-16, 2006.
- [4] J. W. Taylor and P. E. McSharry, "Short-term load forecasting methods: an evaluation based on European data," *IEEE Transactions on Power Systems*, vol. 22, pp. 2213-2219, 2007.
- [5] J. W. Taylor, "Short-term load forecasting with exponentially weighted methods," *IEEE Transactions on Power Systems*, vol. 27, pp. 458-464, 2012.
- [6] L. J. Soares and M. C. Medeiros, "Modeling and forecasting short-term electricity load: a comparison of methods with an application to Brazilian data," *International Journal of Forecasting*, vol. 24, pp. 630-644, 2008.
- [7] S. Fan and R. J. Hyndman, "Short-Term Load Forecasting Based on a Semi-Parametric Additive Model," *IEEE Transactions on Power Systems*, vol. 27, pp. 134-141, Feb 2012.
- [8] A. J. R. Reis, A. P. Alvis, and P. A. d. Silva, "Feature extraction via multiresolution analysis for short-term load forecasting," *IEEE Transactions on Power Systems*, vol. 20, pp. 189-198, 2005.
- [9] Y. Chen, P. B. Luh, C. Guan, Y. Zhao, L. D. Michel, and M. A. Coolbeth, "Short-term load forecasting: similar day-based wavelet neural network," *IEEE Transactions on Power Systems*, vol. 25, pp. 322-330, 2010.
- [10] T. Seniyu, H. Takara, K. Uezato, and T. Funabashi, "One-hour-ahead load forecasting using neural network," *IEEE Transactions Power Systems*, vol. 17, pp. 113-118, 2002.
- [11] A. Troncoso, J. M. R. Santos, J. C. Riquelme, A. G. Expositp, and J. L. M. Ramos, "Time-series prediction: application to the short-term electric energy demand," in *LNAI* vol. 3040, Springer, 2004, pp. 577-586.
- [12] AEMO. Australian Energy Market Operator. Available: www.aemo.com.au
- [13] OMEL. Spanish Electricity Price Market Operator. Available: http://www.omelholding.es/omel-holding/
- [14] J. W. Taylor, "An evaluation of methods for very short-term load forecasting using minite-by-minute British data," *International Journal* of Forecasting, vol. 24, pp. 645-658, 2008.
- [15] I. Koprinska, R. Sood, and V. G. Agelidis, "Variable Selection for Five-Minute Ahead Electricity Load Forecasting," in *Proc. 20th International Conference on Pattern Recognition (ICPR'10)*, Istanbul, Turkey, 2010.
- [16] I. Koprinska, M. Rana, and V. G. Agelidis, "Yearly and Seasonal Models for Electricity Load Forecasting," in *Proc. International Joint Conference on Neural Networks (IJCNN 2011)*, San Jose, USA, 2011.
- [17] A. Kraskov, H. Stögbauer, and P. Grassberger, "Physical Review E," Estimating Mutual Information, vol. 69, 2004.
- [18] H. Yu and B. M. Wilamowski, "Levenberg–Marquardt Training," in Industrial Electronics Handbook. vol. 5 – Intelligent Systems, CRC Press, 2011, pp. 12-1--12-15.