Random Forests Model for One Day Ahead Load Forecasting

Ali Lahouar and Jaleleddine Ben Hadj Slama
Advanced Systems in Electrical Engineering
National Engineering School of Sousse, BP 264, Sousse Erriadh 4023, University of Sousse, Tunisia
National Engineering School of Tunis, BP 37, Le Belvedere 1002 Tunis, Tunisia
Email: ali.lahouar@gmail.com, bhslama@yahoo.fr

***Abstract*—Short term load forecasting is one of the most important tasks for power suppliers, and it is getting more important with deregulation of electricity market and emergence of smart grids. This paper proposes a load prediction model of one day ahead with resolution of one hour, using regression random forests. With information about season, temperature, type of the day and hourly load, a training process is performed to build the adopted model. A real load data set from Tunisian Power Company is used for test, and special attention is paid to the load profile which is specific to warm countries with excessive and unstable demand in summer. The results reflect accuracy and effectiveness of the proposed method, keeping low prediction error for long test periods.**

***Keywords*—Short term load forecasting, artificial intelligence, random forest, smart grid.**

I. Introduction

Load forecasting is predicting how the electric load demand waveform of an area will look like for a certain period of time in the future, using a mathematical model. It is one of the old and regenerate issues in the literature and scientific community. Following the evolution of power installations, it was a research topic for decades, and has encountered since the beginning of the 21th century new challenges related to market deregulation and appearance of smart grids.

In a deregulated electricity market where consumers can choose their suppliers and where the price is not fixed, load forecasting is of a great importance for both sides. The supplier has the possibility to estimate the demand of his customers and thus optimize his production and reduce its costs, and the consumer can make use of electricity at low cost, or even plan his consumption according to the price change. More advantages are mentioned in literature, such as load dispatch decisions and maintenance planning [1], [2]. Although deregulation is dominating the market in many countries, this is not the case in Tunisia, where there is only one supplier, and where the prices are regulated by the government. But still, forecasting the load is very important due to emergence of renewable energy in a smart grid environment. In such conditions, innovative computing technologies are utilized to optimize energy production and distribution. Among these technologies, forecasting the load demand is featured and should become more accurate. This is justified considering the intermittence of renewable energy, which may cause unexpected voltage variations apart from increasing forecast error [3]. An accurate and robust forecast may result in better production planning and smarter management of classic thermal power plants. In Tunisia, only wind energy pushes towards sophisticating prediction models because photovoltaic energy has no effective contribution in the grid, and hydroelectricity does not have problem of intermittence.

Load forecast may be achieved for short, medium and long term. Even if there is no explicit explanation of what each time horizon is, it can be assumed that short term ranges from one hour to seven days of prediction ahead, medium term from one week to several months ahead and long term for more than one year ahead. Each type has its properties and applications, and the short term forecast, which is the most found in literature, remains a very hard task [2]. Speaking about prediction models, they arrived to maturity after decades of development. They can be divided into two main groups; classic statistical methods and artificial intelligence methods. Classic methods establish a strict mathematical relationship between inputs and outputs, such as multiple linear regression [4], exponential smoothing [5], autoregressive moving average (ARIMA) [6] and Kalman filtering [7], [8]. These models are criticized for their inability to handle non-linear problems [9]. Artificial intelligence methods in general are able to predict an output after a training process. The most commonly used methods in this category are artificial neural networks (ANN) [2], [10]–[17] and then support vector machine (SVM) [1], [18]–[22]. A class of ANN called self-organizing map or Kohonen network are also used to this end [1], [18]. These models are powerful and able to incorporate non-linearity, but have their own drawbacks, such as local optimum trap in ANN, and parameter tuning in SVM [9]. Expert systems based on fuzzy logic are a good alternative and provide results interpretability [23], [24], but have also their weakness points [18]. In addition, other less-known methods are applied such as random forests (RF) [25], grey prediction theory [26], [27] and Gaussian process [3]. The numerous drawbacks of each model push the researchers to use hybrid methods, especially after appearance of evolutionary optimization algorithms such as differential evolution (DE), genetic algorithms (GA) and particle swarm optimization (PSO). It is possible for example to use DE to tune SVM parameters [9], combine fuzzy inference and PSO [28] or use chaotic feature selection [29].

This paper tries to focus on random forests and their use to solve regression problem. This method did not receive much attention in load forecasting despite its accuracy, rapidity and

978-1-4799-7947-9/15/$31.00 © 2015 IEEE

especially easiness to choose its parameters. It is able to give results as good as optimized ANN and SVM without need to be tuned, and even a bad choice of its parameters does not affect drastically the prediction accuracy. Random forest is used to construct a generic model able to predict the load demand of one day ahead by a step of one hour. This model should give accurate results whatever the day or the season is, and for long test periods. The rest of the paper is organized as follows: section 2 explains briefly the mathematical structure of random forests, section 3 details the adopted model, section 4 gives the load profile characteristics, simulation results and comparisons and section 5 concludes the paper.

II. Mathematical preliminaries

A brief explanation of random forests is given in this section. For full mathematical development, see references [30] and [31].

A. Decision tree

A decision tree is a statistical model for classification and regression introduced by Breiman in 1984. Forecasting is a regression problem; hence regression trees are used here. Like other machine learning techniques, it requires a phase for training and a phase for test. Lets consider a vector X of n features as input, a corresponding label Y as output and a training set S of m couples (X, Y).

$$X = (x_1, ..., x_j, ..., x_n)^T, X \in \mathbb{R}^n \tag{1}$$

$$Y \in \mathbb{R} \tag{2}$$

$$S = \{(X_1, Y_1), ..., (X_m, Y_m)\} \tag{3}$$

The training process consists in building a predictor h by partitioning the features recursively into nodes with different labels Y until a certain termination criterion is met. This criterion is generally when it is not possible to have children nodes with different labels. The terminal nodes are called leaves of the tree and they represent the different possible labels Y. Then, the training phase consists in predicting the label of any new feature vector using the build predictor h.

$$Y = h(X) \tag{4}$$

B. Random forest

The random forest is generalization of decision trees proposed by Breiman in 2001. It is an ensemble method which combines prediction of weak predictors h_i. The most important parameters are number of trees $ntree$ and number of variables to partition at each node $mtry$.

$$Y = h(X) = \frac{1}{ntree} \sum_{i=1}^{ntree} h_i(X) \tag{5}$$

III. The adopted model

The model is constructed by 24 random forest predictors standing for the 24 hours of the day. The choice of predicting each hour independently is made since the load at a certain hour is really close to the load at the same hour of the previous day, and not to the previous hour. For the load predictor of the day d at hour h, the features of the input vectors are the factors affecting the load demand. These factors are month number, type of the day, minimum and maximum temperature of day d, load of day $d-1$ at hour h, morning and evening load peak of day $d-1$. The output (label) is the load of day d at hour h. Hence, the whole model is able to predict the 24 coming hours of load demand as shown in Fig. 1.

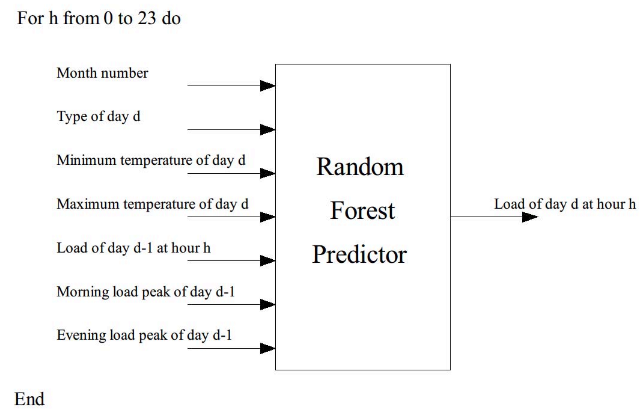

Fig. 1. Forecasting Model

There is no general rule for choosing inputs. In literature, some factors are known to drive electricity consumption, and may vary according to the case. In this work, the chosen factors are those which give the minimum prediction error, without any optimization process. The temperature, considered as the most influencing variable [3], is measured at Tunis, the capital and the most consuming region in the country. The temperature forecast of day d, which is among inputs, should be as precise as possible; otherwise the accuracy of the predictor will be affected.

IV. Case study

A. The load profile

The available data are half-hourly load demand of the whole Tunisia in 2009 and 2010. As shown in Fig. 2, a summer day is obviously superior to days of other seasons because of air conditioners. Furthermore, all curves have almost the same shape with small shifts, which is favorable for predictors. The period of work between 8 am and 5 pm is characterized by stable demand except for summer, and conserves this stability in Sundays and holidays but with less demand. A peak appears between 5 pm and 9 pm according to seasons.

The load demand is increasing every year, and this is due to many factors. The increasing number of sold household appliances especially air conditioners is the main reason. This growth may be estimated thanks to sale reports. However, smuggling of electrical devices, which means import of unreported goods, affects the load and makes the forecast harder due to lack of statistics. Population growth and urbanization ratio result in greater number of energy consumers and contribute also to this long term demand raise. Although photovoltaic energy is not yet fully integrated in the Tunisian grid, the trade of solar panels for water heating is going well in the five recent years, and this should be taken into account despite its slight impact. Solar panels ensure heating water for free for at least

6 months in the year and may lighten the load in the summer. All these factors have their impact only for the long term and they are not included in the prediction model, but they explain the yearly evolution of the load profile and why it is hard to predict the summer load behavior.

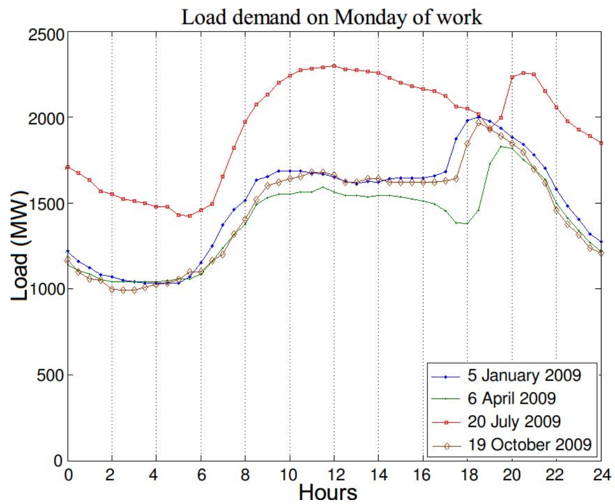

Fig. 2. Daily load profile

B. Results

As a first step, training is performed only one time for the whole test period. For example, to predict the daily load for all the days of January 2010, it is enough to train one time the machine with the data of the year 2009. In other words, there is no need to perform new training every day. This approach has the advantage of being very fast and adapted for long test period. The mean average percentage error MAPE defined in equation (6) is the most common criterion to evaluate prediction accuracy, where $L(h)$ and $\hat{L}(h)$ are measured and predicted load at hour h, respectively. This error is computed every day and averaged at the end of test period.

$$MAPE = (\frac{1}{24}\sum_{h=0}^{23}\frac{|L(h)-\hat{L}(h)|}{L(h)}) \times 100 \qquad (6)$$

Table I shows the monthly average MAPE of four test months in 2010, and the time taken by the calculus loop in a computer equipped with Intel i7 processor. These results are acceptable except for July where the error is high. Fig. 3 gives the worst prediction case which caused this error peak, where an unexpected load behavior occurs; three successive days of extremely high load demand followed by holiday. This high load exceeding 3000 MW is due to high temperature and excessive air conditioning, and was not present in the training data, which explains the machine inability to predict these values. Having larger training sets requires more calculus time and does not necessarily improve forecasting accuracy; it may even result in worse results. A suitable choice is to select a training period great enough to learn all possible cases and close to test period.

Fig. 4 shows the daily MAPE of January 2010 for RF, ANN and SVM with the configuration of Table II, all of them using the same forecasting model. It is important to mention

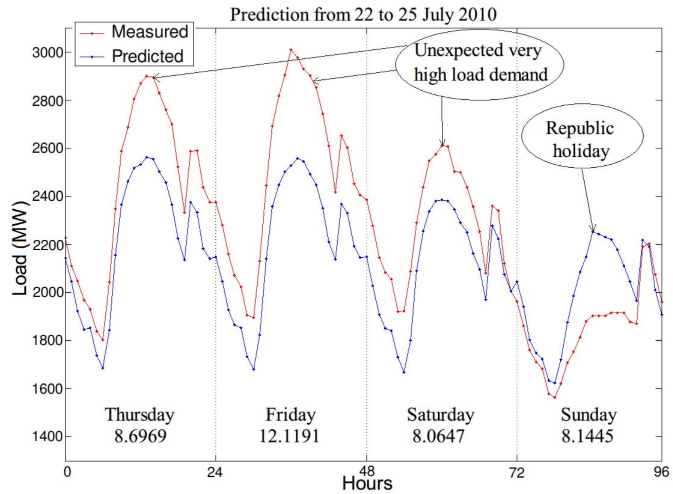

Fig. 3. Worst prediction case with the daily MAPE

Training period	Test period	Monthly MAPE (%)	Calculus time (s)
January 2009 - December 2009	January 2010	2.5656	1.28
January 2009 - March 2010	April 2010	2.1903	1.70
January 2009 - June 2010	July 2010	4.0202	2.16
January 2009 - September 2010	October 2010	2.2336	2.68

TABLE I. FORECASTING RESULTS

that the parameters are chosen manually with no optimization process, and that ANN and SVM may give much more accurate results with suitable parameters, but may also take much more time to be executed. The time taken by the calculus loop is not significant for one day ahead forecast since very fast computing is not required. It is given only to compare methods and show the impact of varying parameters on accuracy and rapidity.

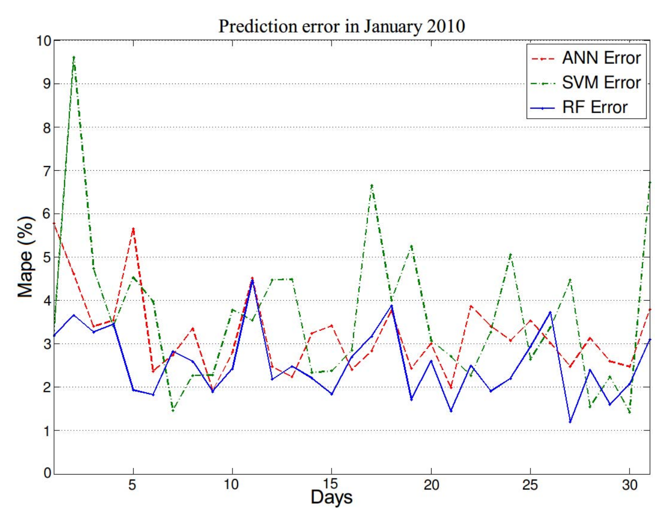

Fig. 4. Comparison between different methods

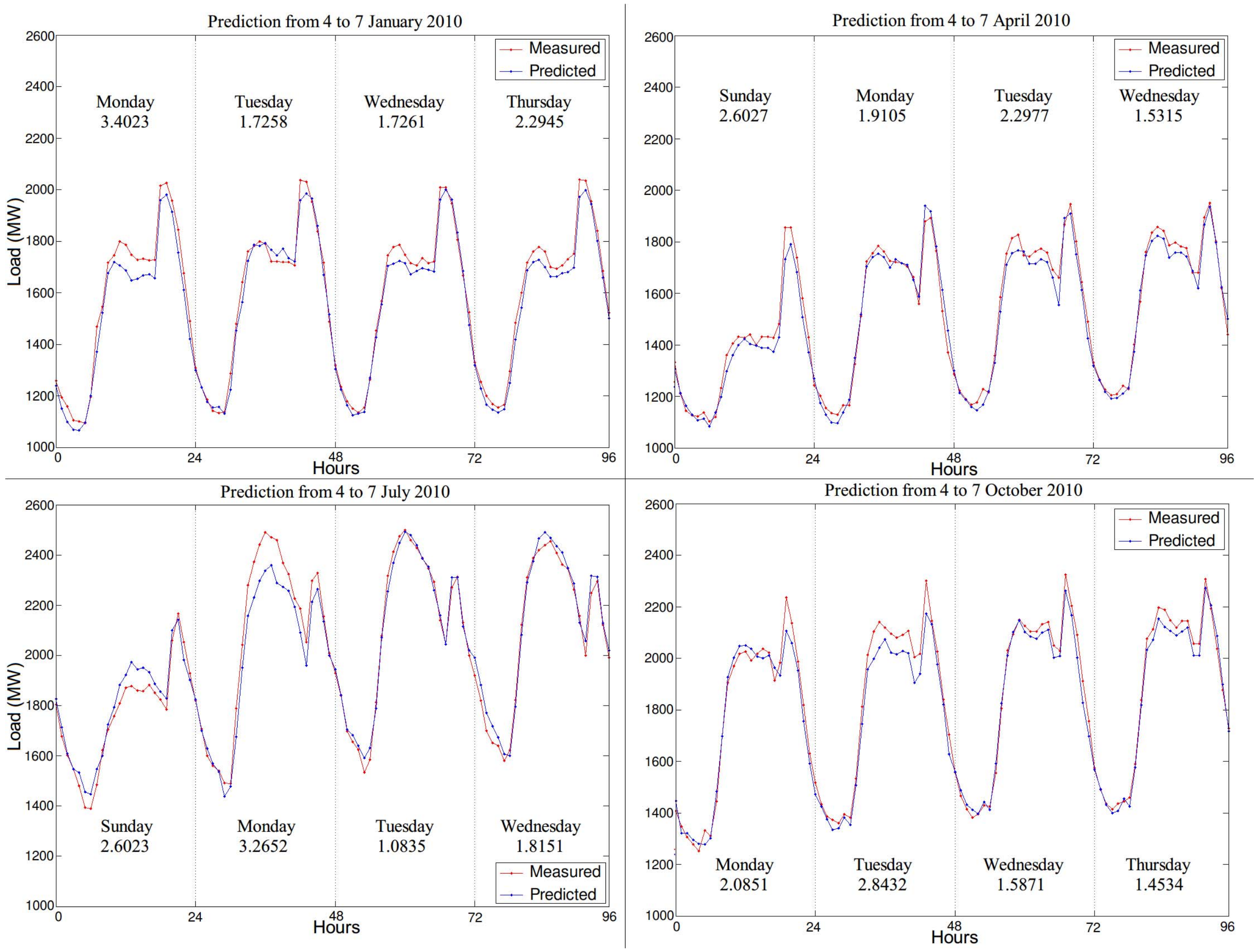

Fig. 5. Prediction results in four seasons with the daily MAPE

Method	Configuration	MAPE (%)	Calculus time (s)
RF	$ntree$=100 $mtry$=2.5	2.5656	1.28
ANN	Feed forward network 1 hidden layer, 10 neurons Levenberg-Marquardt algorithm Sigmoid activation	3.2641	12.46
SVM	Polynomial kernel Degree 1 Regularization C=1 γ=0.0003	3.6765	0.44

TABLE II. COMPARISON BETWEEN METHODS

Test period	Monthly MAPE (%)	Calculus time (s)
January 2010	2.2524	40.40
April 2010	2.1569	52.03
July 2010	3.1416	67.70
October 2010	2.2092	83.25

TABLE III. RESULTS OF MOVING TRAINING PERIOD

C. Improvements

As a second step, the training process is performed every day. Hence, the training period is moving and the real load value of each day is used to train the machine again and predict the next day. As a consequence, the calculus loop took much more time, but an improvement is visible, especially in the case of July. Table III summarizes the new found results, and Fig. 5 shows the differences between measured and predicted curves in four days of four different seasons. The calculus time in every month is growing since the training period is growing too.

Fig. 6 shows much more accurate forecasting results than before, where the prediction curve is able to reach the new values that were learned by the machine. However, the prediction of the holiday is not precise. The load fall is underestimated due the conditions of being Sunday and holiday at the same time in addition to high demand of previous days. Contrariwise, Fig. 7 indicates better estimation of the load fall in the holiday of 26 February. In fact, predicting holidays is a hard task, since they are rare and do not allow the machine to learn their behavior correctly. Among possible solutions for an eventual future work, holidays may have the same day type of Sundays, since they are similar. A manual post-processing adjust may also be performed according to the estimated temperature.

The MAPE of 2% is similar to the error found in recent researches for one day ahead prediction, which ranges between 1% and 2% [3], [28], [29]. The proposed approach has the advantage to perform tests for long periods without losing this low error. In addition, it is very fast and does not require any optimization process like other methods.

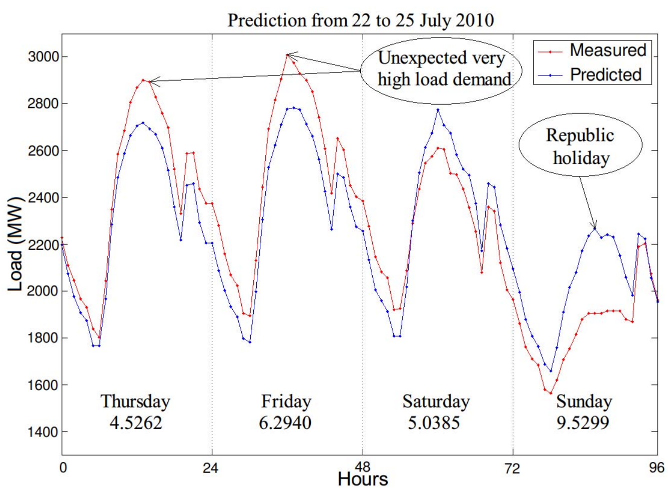

Fig. 6. Solution to the worst prediction case

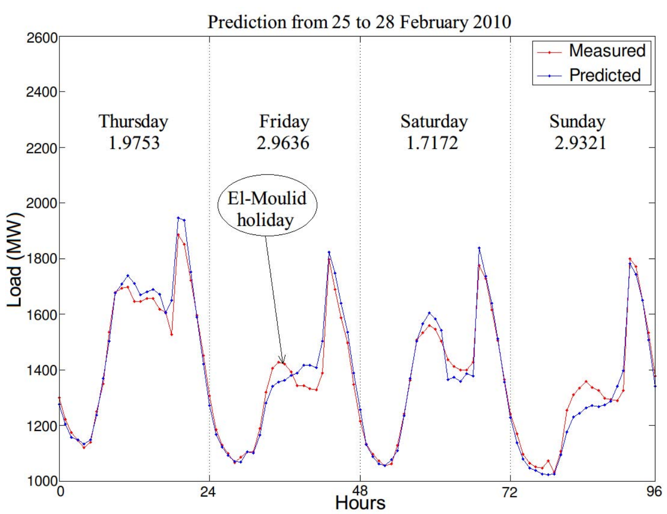

Fig. 7. Predicting a holiday

V. Conclusion

The electric load estimation of one day ahead is critical for both power suppliers and customers in smart grid environment. This paper proposed a novel prediction model for short term load forecasting based on random forests. This method is characterized by rapidity and easiness of parameters tuning compared to other artificial intelligence methods, namely ANN and SVM. Real data from Tunisian power grid is used to test the prediction machine for long periods up to one month. The results feature low prediction error especially when training is repeated every day. However, rare exceptions are found for some special days with unexpected behavior, where a manual intervention is required to adjust the results.

Acknowledgment

The authors would like to thank the Tunisian Company of Electricity and Gas (STEG) for providing half-hourly load data which were used in this work.

References

[1] J. Nagi, K. S. Yap, F. Nagi, S. K. Tiong, and S. K. Ahmed, "A computational intelligence scheme for the prediction of the daily peak load," *Applied Soft Computing*, vol. 11, no. 8, pp. 4773 – 4788, 2011.

[2] S. Kulkarni, S. P. Simon, and K. Sundareswaran, "A spiking neural network (snn) forecast engine for short-term electrical load forecasting," *Applied Soft Computing*, vol. 13, no. 8, pp. 3628 – 3635, 2013.

[3] H. Mori and E. Kurata, "An efficient kernel machine technique for short-term load forecasting under smart grid environment," in *Power and Energy Society General Meeting, 2012 IEEE*, July 2012, pp. 1–4.

[4] N. Amral, C. Ozveren, and D. King, "Short term load forecasting using multiple linear regression," in *Universities Power Engineering Conference, 2007. UPEC 2007. 42nd International*, Sept 2007, pp. 1192–1198.

[5] J. W. Taylor, "An evaluation of methods for very short-term load forecasting using minute-by-minute british data," *International Journal of Forecasting*, vol. 24, no. 4, pp. 645 – 658, 2008, energy Forecasting.

[6] M. Cho, J. Hwang, and C.-S. Chen, "Customer short term load forecasting by using arima transfer function model," in *Energy Management and Power Delivery, 1995. Proceedings of EMPD '95., 1995 International Conference on*, vol. 1, Nov 1995, pp. 317–322 vol.1.

[7] H. Al-Hamadi and S. Soliman, "Short-term electric load forecasting based on kalman filtering algorithm with moving window weather and load model," *Electric Power Systems Research*, vol. 68, no. 1, pp. 47 – 59, 2004.

[8] F. Zhao and H. Su, "Short-term load forecasting using kalman filter and elman neural network," in *Industrial Electronics and Applications, 2007. ICIEA 2007. 2nd IEEE Conference on*, May 2007, pp. 1043–1047.

[9] J. Wang, L. Li, D. Niu, and Z. Tan, "An annual load forecasting model based on support vector regression with differential evolution algorithm," *Applied Energy*, vol. 94, no. 0, pp. 65 – 70, 2012.

[10] N. Kandil, R. Wamkeue, M. Saad, and S. Georges, "An efficient approach for shorterm load forecasting using artificial neural networks," in *Industrial Electronics, 2006 IEEE International Symposium on*, vol. 3, July 2006, pp. 1928–1932.

[11] A. Bakirtzis, V. Petridis, S. Kiartzis, M. Alexiadis, and A. Maissis, "A neural network short term load forecasting model for the greek power system," *Power Systems, IEEE Transactions on*, vol. 11, no. 2, pp. 858–863, May 1996.

[12] S. Kouhi and F. Keynia, "A new cascade {NN} based method to short-term load forecast in deregulated electricity market," *Energy Conversion and Management*, vol. 71, no. 0, pp. 76 – 83, 2013.

[13] A. Azadeh, S. Ghaderi, and S. Sohrabkhani, "Annual electricity consumption forecasting by neural network in high energy consuming industrial sectors," *Energy Conversion and Management*, vol. 49, no. 8, pp. 2272 – 2278, 2008.

[14] T. Iizaka, T. Matsui, and Y. Fukuyama, "A novel daily peak load forecasting method using analyzable structured neural network," in *Transmission and Distribution Conference and Exhibition 2002: Asia Pacific. IEEE/PES*, vol. 1, Oct 2002, pp. 394–399 vol.1.

[15] J. cong Cao, "Application of {ELD} and load forecast in optimal operation of industrial boiler plants equipped with thermal stores," *Applied Thermal Engineering*, vol. 27, no. 23, pp. 665 – 673, 2007.

[16] W. Dai and P. Wang, "Application of pattern recognition and artificial neural network to load forecasting in electric power system," in *Natural Computation, 2007. ICNC 2007. Third International Conference on*, vol. 1, Aug 2007, pp. 381–385.

[17] H.-T. Zhang, F.-Y. Xu, and L. Zhou, "Artificial neural network for load forecasting in smart grid," in *Machine Learning and Cybernetics (ICMLC), 2010 International Conference on*, vol. 6, July 2010, pp. 3200–3205.

[18] J. Che, J. Wang, and G. Wang, "An adaptive fuzzy combination model based on self-organizing map and support vector regression for electric load forecasting," *Energy*, vol. 37, no. 1, pp. 657 – 664, 2012, 7th Biennial International Workshop Advances in Energy Studies.

[19] C.-T. Lin and L.-D. Chou, "A novel economy reflecting short-term load forecasting approach," *Energy Conversion and Management*, vol. 65, no. 0, pp. 331 – 342, 2013, global Conference on Renewable energy and Energy Efficiency for Desert Regions 2011.

[20] C.-H. Wu, G.-H. Tzeng, and R.-H. Lin, "A novel hybrid genetic algorithm for kernel function and parameter optimization in support vector regression," *Expert Systems with Applications*, vol. 36, no. 3, Part 1, pp. 4725 – 4735, 2009.

[21] W.-C. Hong, "Application of chaotic ant swarm optimization in electric load forecasting," *Energy Policy*, vol. 38, no. 10, pp. 5830 – 5839, 2010, the socio-economic transition towards a hydrogen economy - findings from European research, with regular papers.

[22] ——, "Chaotic particle swarm optimization algorithm in a support vector regression electric load forecasting model," *Energy Conversion and Management*, vol. 50, no. 1, pp. 105 – 117, 2009.

[23] S. Kiartzis, A. Bakirtzis, J. Theocharis, and G. Tsagas, "A fuzzy expert system for peak load forecasting. application to the greek power system," in *Electrotechnical Conference, 2000. MELECON 2000. 10th Mediterranean*, vol. 3, May 2000, pp. 1097–1100 vol.3.

[24] S. Papadakis, J. Theocharis, S. Kiartzis, and A. Bakirtzis, "A novel approach to short-term load forecasting using fuzzy neural networks," *Power Systems, IEEE Transactions on*, vol. 13, no. 2, pp. 480–492, May 1998.

[25] Y.-Y. Cheng, P. Chan, and Z.-W. Qiu, "Random forest based ensemble system for short term load forecasting," in *Machine Learning and Cybernetics (ICMLC), 2012 International Conference on*, vol. 1, July 2012, pp. 52–56.

[26] G. Zheng, S. Chen, and S. Fan, "A power load forecasting method based on matching coefficient of meteorological factors and similar load modification," in *Intelligent Human-Machine Systems and Cybernetics (IHMSC), 2012 4th International Conference on*, vol. 1, Aug 2012, pp. 216–219.

[27] N. Tang and D.-J. Zhang, "Application of a load forecasting model based on improved grey neural network in the smart grid," *Energy Procedia*, vol. 12, no. 0, pp. 180 – 184, 2011, the Proceedings of International Conference on Smart Grid and Clean Energy Technologies (ICSGCE 2011.

[28] X. Yang, J. Yuan, J. Yuan, and H. Mao, "An improved {WM} method based on {PSO} for electric load forecasting," *Expert Systems with Applications*, vol. 37, no. 12, pp. 8036 – 8041, 2010.

[29] S. Kouhi, F. Keynia, and S. N. Ravadanegh, "A new short-term load forecast method based on neuro-evolutionary algorithm and chaotic feature selection," *International Journal of Electrical Power & Energy Systems*, vol. 62, no. 0, pp. 862 – 867, 2014.

[30] L. Breiman *et al.*, *Classification and Regression Trees*. New York: Chapman & Hall, 1984.

[31] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.