Electric Vehicle Load Forecasting using Data Mining Methods

Conference Paper · November 2013

DOI: 10.1049/cp.2013.1914

Electric Vehicle Load Forecasting using Data Mining Methods

***E S Xydas*[1], *C E Marmaras*[2], *L M Cipcigan*[3], *A S Hassan*[4], *N Jenkins*[5]**

[1]Student Member IET, UK, XydasE@cardiff.ac.uk, [2]Member IEEE, UK, MarmarasC@cardiff.ac.uk,
[3]Member IET Lecturer, UK, CipciganLM@cardiff.ac.uk, [4]Member IET, UK, SaniHassanA@cardiff.ac.uk,
[5]Fellow IET Professor, UK, JenkinksN6@cardiff.ac.uk

Keywords: Electric Vehicle, Forecast, Data Mining.

Abstract

The continuous growth and evolve of vehicle electrification causes the electric power systems to confront new challenges, since the load profile changes, and new parameters are being set. With the number of EVs gradually rising, problems may occur in technical characteristics of the network, like bus voltages and line congestion [1]. Therefore, it is necessary to develop EV management systems so as to prevent such phenomena. The effectiveness of such systems is heavily depended on the early knowledge of future demand. This knowledge can be provided by accurate EV load forecasting techniques. In this paper, the use of various data mining methods is examined and their performance in EV load forecasting is evaluated.

1 Introduction

According to the U.K. Department for Transport [2], road transport is a significant contributor to greenhouse emissions and reductions are required for moving U.K. to a low carbon future in order to meet the Climate Change Act targets. The electrification of transport offers a good opportunity to decrease CO_2 emissions and increase the national energy security. The financial incentives, in combination with the increasing prices in oil, lead customers to consider electric vehicles (EVs) as a reliable and economical solution for transportation.

The development of an EV market is strongly dependent on the parallel development of the recharging infrastructure which will result in a spatially uneven increase in the electricity demand. The U.K. government supports the penetration of ultra-low emission vehicles by announcing recently a £37 million funding package for providing 75% of the cost of installing new charging points in order to motivate the EVs drivers to reduce their range anxiety.

Electric vehicles are a mobile source of demand, charged for relatively long periods of time and as a result of this, EVs could place significant coincident demand on the system. According to [3], if all the registered vehicles in United States had to charge 5-10kWh on a daily basis, this would lead to an increase of 12-23% at the electricity generation requirement. For UK, an uncontrolled EV charging regime increases the British winter day peak demand by 3.2 GW (3.1%) for a low EV uptake case (7%) and the British winter day peak demand by 37GW (59.6%) for a high EV uptake case (48.5%), for the year 2030 [4, 5].

The uncontrolled charging of EVs might increase the system's peak demand, exceeding voltage limits and/or overloading lines and transformers [6-11]. According to [9], when higher level of EV penetration is considered, such phenomena are more often and intense. In order to prevent grid technical violation and avoid early reinforcements of existing infrastructure, there is a need for coordinating EV charging. In order for this coordination to be effective, estimation of the day-ahead charging demand should be taken into consideration. This estimation could be provided from EV load forecasting methods.

EV load forecasting is influenced by fluctuating factors like driving and travel patterns of each EV owner. These patterns are important to be considered in order to approximate the charging demand. The stochastic nature of the EV charging demand factors forces the use of advanced forecasting techniques which are able to decipher all the patterns. The use of artificial intelligent techniques enables the decoding of complicated historical charging events despite the high randomness they appear.

The scope of this paper is to evaluate the performance of various artificial intelligence techniques in forecasting the charging demand of an EV fleet. For this evaluation, two different case studies were considered based on real electric vehicles charging events.

Section 2 briefly describes four data mining methods used for forecasting the EV demand. In Section 3 are described the case studies for evaluating the accuracy of the data mining methods. Finally conclusions are drawn in Section 4.

2 Data Mining methods

Data Mining refers to extracting or mining knowledge from large amounts of data [12]. This extraction reveals hidden relationships such as patterns and association among features in datasets. This "filtering" of the unnecessary information is important, as it saves time and space in the modern systems where effectiveness is measured in such values. Data mining is also known as Knowledge Discovery from Databases (KDD), and refers as the non-trivial extraction of implicit, previously unknown and potentially useful information from databases. The KDD procedure includes also a pre-processing stage where missing or incorrect data are either removed or corrected. |In this process normalization of the data can be performed. After the completion of this stage, the formatted data are forwarded to the core of KDD process, the data mining. In this step intelligent techniques are applied to interpret the hidden patters behind the data. This procedure is illustrated in Figure 1.

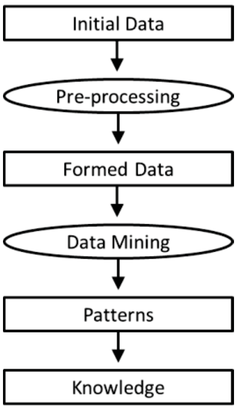

Figure 1: Flowchart of the KDD process

In this paper four data mining methods were considered, described in the following subsections.

2.1 Decision Tables

Decision Table algorithms build and use a simple decision table majority classifier as proposed by Kohavi [17]. The dataset is summarized with a decision table which contains the same number of attributes as the original dataset. The simplicity of creating and reading a decision table is one of the method's main advantages. The main structure of a decision table is shown in Table 1.

	Decision Rule 1	Decision Rule 2	Decision Rule 3
Conditions		Condition Entries	
Actions		Action Entries	

Table 1: General structure of a Decision Table [18]

In general, a decision table is divided into four quadrants. The upper two quadrants contain the conditions for each Decision rule. The lower two quadrants describe all the possible actions for every corresponding condition. However, one important drawback is that in complex datasets with many attributes Decision tables may became extremely large.

2.2 Decision Trees

Decision Trees are a supervised learning method used for classification tasks. Decision Trees are used to classify instances by categorizing them taking into account the feature values. All the middle nodes represent an evaluation of a condition (Decision Nodes) and the terminal nodes (Leaf Nodes) represent the decision result. A typical structure of a decision tree is shown in Figure 2.

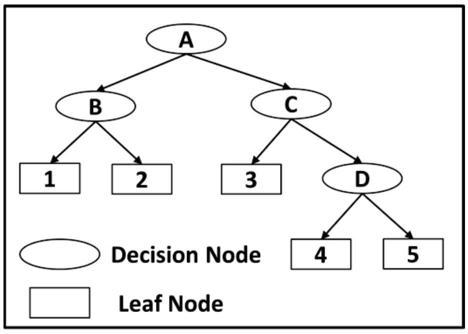

Figure 2: Typical structure of a decision tree

In this paper the Reduced Error Pruning (REP) Tree was used. REP Tree algorithm is a fast decision tree learner [19]. It builds a decision tree using known information and prunes it using reduced-error pruning. With this algorithm the possibility of pruning sub-trees is examined and evaluated according to the reduction or not of the error. In case of an error reduction, the sub-tree is pruned and the final tree is smaller and more accurate.

2.3Artificial Neural Networks (ANN)

Artificial Neural Networks (ANN) was inspired from the human neurons. ANN are able to find hidden correlations between input data and target value sand solve complicated problems despite noise and fluctuation in the data. There are various types of ANN and the most known are the Multi-layered Perceptron (MLP), Radial Basis Function (RBF) and the Kohonen networks [20]. In this research MLP was selected to provide forecasts. MLP consist of three basic layers: The Input Layer, the Hidden Layer and the Output Layer [12]. The Input layer can have any number of inputs. The Hidden Layer can contain one or more (sub) layers and each of them can contain one or more nodes. They are called "Hidden" because they receive internal inputs and produce internal outputs, not directly connected to the external layers. The only existing connections are between the input layer and first hidden layer, and the last hidden layer and the output layer. The structure of an MLP neural network is described in Figure 3.

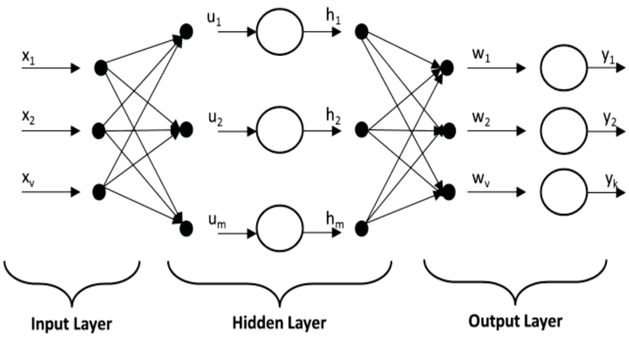

Figure 3: Structure of an MLP neural network

2.4 Support Vector Machines (SVM)

Support Vector Machines (SVM) is a machine learning method associated with classification, regression and other learning tasks and was developed by Boser, Guyon and Vapnic [21].

SVM try to find linear separations between the data ("decision boundaries" for separating one class from another) [12]. Assuming data with two attributes, SVM depict them into a two dimensional space and search for possible separating lines. If the data are depended on three attributes, they are projected on a three dimensional space, and SVM search for the possible separating planes. Generalizing for n-attributes, the depiction is on an n-dimension space and SVM search for separating hyperplanes. SVM will find many different lines or hyperplanes which divide the data. The optimal line/hyperplane is selected based on the maximization of the separating distance.

A simple two-dimensional case is illustrated in Figure 4.

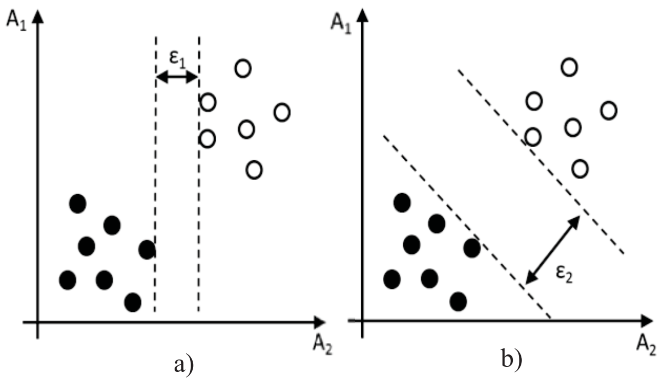

Figure 4: a) A random pair of separating lines, b) The pair of separating lines with the maximum distance

In case a linear separation is not possible, SVM transform the data into higher dimensions using nonlinear mapping. Then, the process of finding the linear hyperplane with the maximum margin in the new space is repeated.

3 Modelling framework

The data mining tasks were conducted on an Intel i3 Processor Platform, which consists of 3GB RAM and Microsoft Windows 7 operating system. WEKA 3.6.9software tool was used [22]. WEKA is a widely known toolkit for machine learning and data mining algorithms such as regression, classification, clustering, association rules, visualization and data processing, developed by the University of Waikato. Data sets used in WEKA was in Comma-Separated Values file format (.CSV), where values are separated by commas and sorted according to the attribute.

The general procedure that was followed is described in Figure 5.

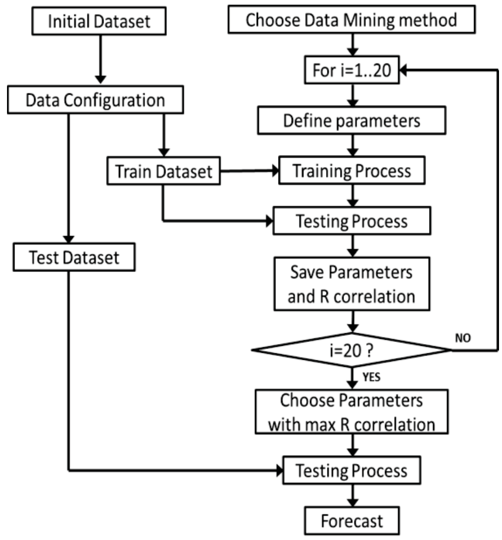

Figure 5: Flowchart of the modelling process

The modelling framework consists of two parts: the data configuration (Pre-processing stage), and the data processing (data mining stage).

In the data configuration stage the initial data set is divided in two parts, namely Train Dataset and Test Dataset. The Train Dataset is used to train the model, and to find hidden correlations and patterns between attributes and target values. In general, the majority of the data consist the Train Dataset. The Test Dataset is used to evaluate the performance of the data mining method. A typical structure of a Train/Test Dataset readable from the model is described in Table 2:

Target Title	**Attribute_1 Title**	…	**Attribute_M Title**
Target Value-1	Attribute_1 Value-1	…	Attribute_M Value-1
Target Value-2	Attribute_1 Value-2	…	Attribute_M Value-2
…	…	…	…
Target Value-N	Attribute_1 Value-N	…	Attribute_M Value-N

Table 2: Structure of a Train/Test Dataset

At the beginning of the data processing stage a data mining method and its parameters are selected. Then, the model is trained and tested on the Train dataset, and its accuracy is evaluated. This train/test sequence is repeated for twenty different combinations of parameters, and the combination that gives the best accuracy is selected. Once the most accurate train model is build, it will be tested on the Test Dataset. In this testing process, the model uses the knowledge obtained by the Train Dataset to provide a forecast for the Target values of the Test Dataset.

The accuracy of the model output was assessed using Mean Absolute Percentage Error (MAPE), Root Mean Square Error

(RMSE) and r-Correlation. The training and testing duration were also considered to evaluate the performance.
The formula for calculating the MAPE, RMSE and r-Correlation are:

$$MAPE = \frac{\sum_{i=1}^{N}\left(\frac{|X_i - Y_i|}{X_i}\right)}{N} \times 100\% \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(X_i - Y_i)^2}{N}} \quad (2)$$

$$r = \frac{\sum_{i=1}^{N}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{N}(X_i - \overline{X})^2}\sqrt{\sum_{i=1}^{N}(Y_i - \overline{Y})^2}} \quad (3)$$

where

N is the number of predicted values,
X is the actual values and
Y is the forecasted values.

3.1 Case Study 1: Residential charging Stations in USA

The EV Project is a large deployment of EVs and charging infrastructure in USA launched by ECOtality on October 1, 2009 [23]. With grants received from US Department of Energy and the support of various Industrial Partners, Electric Vehicle Supply Equipment (EVSE) was installed in major cities and metropolitan areas across the United States. By the end of 2012 7,376 electric vehicles (Nissan Leafs, Chevrolet Volts and Smart4Two) participated in the project [23]. A total number of 9,333 EVSE were installed, 6,694 of which are residential, 2,583 commercial and 56 DC fast chargers. For this case study, aggregated residential data from the 4th Quarter of 2012 were provided by ECOtality, in order to test the performance of various data mining methods.

3.1.1 Data Configuration

The data from ECOtality project include distribution curves that were used to create a representative EV fleet and its charging demand for one year. Figures 6 and 7 show the distribution of the energy consumption and the duration per charging event.

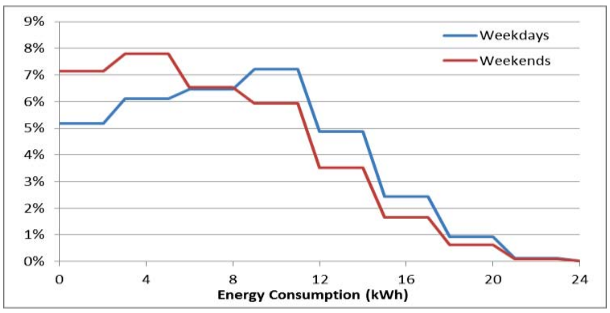

Figure 6: Energy consumption distribution / charging event

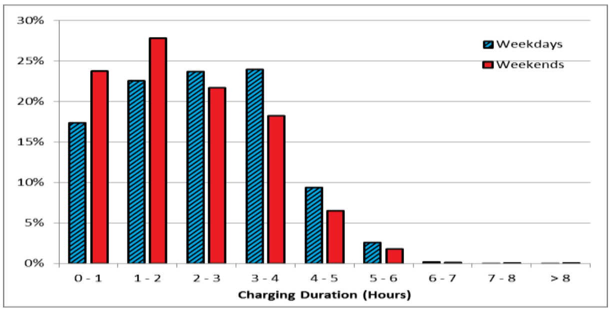

Figure 7: Charging duration distribution / charging event

A fleet of 3000 EVs was considered and their charging events were created based on the above distributions. Arrival times for the EVs were estimated using the Charging Availability graph for a typical weekday/weekend as shown in Figure 8 [23].

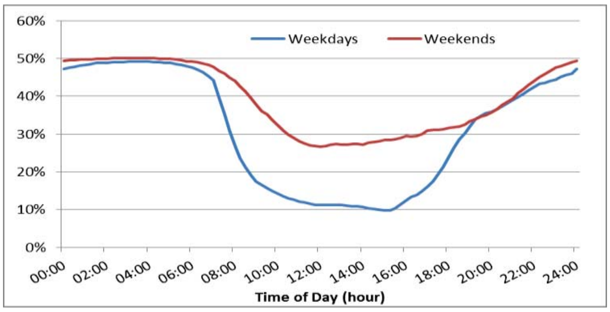

Figure 8: Charging Availability for 24 hours

Using the above data, the aggregated charging demand was created for one year on a half-hourly basis. The attributes for each half hour are:

1. *Previous Day Load*: the charging demand of previous day for each half hour.
2. *Week:* number of the week (1-53).
3. *Day:* number of the day (1-7) starting with Monday.
4. *Type of Day:* Weekday or Weekend.
5. *Half Hour:* 1-48 half hour parts of each day.
6. *Number of New Connections:* the new EV plug-in connections for every half hour.
7. *Total Charging Connections:* the number of EV that are connected and charging for every half hour.

Once the dataset is structured using the template presented in Table 2, the last day of the dataset was selected for the Test Dataset. All the previous ones are the Train Dataset.

3.1.2 Results

Using the procedure described in Figure 5, the forecasts of the four data mining methods are shown in Figure 9 in comparison with the actual demand.

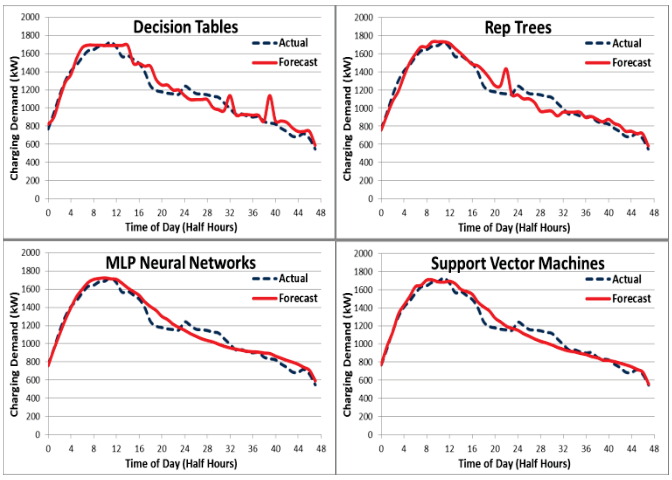

Figure 9: Charging Demand Forecasts

Performance measures for each method are summarized in Table 3.

Method / Index	Decision Tables	Rep Trees	MLP ANN	SVM
MAPE (%)	6.407	5.675	5.387	4.616
RMSE	87.73	83.99	70.92	67.05
r-Correlation (%)	96.69	96.83	97.84	98.09
Training Time (s)	0.7	0.22	26.71	48.89
Testing Time (s)	0.64	0.25	27.01	48.25

Table 3: Performance measures for each method

The results show that SVM provided the most accurate forecasts, but the training process was slow compared to the other methods. On the other hand, ANN performed well without being so time consuming. The other two methods, despite being less accurate, they need the least time to produce a trained model. The MAPE criterion is the accepted industry standard for measuring load forecasting accuracy [24] and a value less than 5% is acceptable [25]. Large absolute errors are penalized strongly in the RMSE index.

3.2 Case Study 2: Public charging stations in France

The data for this case study consist of real charging events occurred in public charging stations. The period that these charging events took place was from April 2011 until February 2012. This was a pilot project in which 71 electric vehicles and their charging activities were monitored.

3.2.1 Data Configuration

The charging events were classified according to the id of each EV, and examined individually. For each EV, charging patterns like the connection/disconnection time and the energy demand per charging event were analysed in order to produce weekly distributions of that characteristics. An example of the charging demand distribution of four random EVs for one week is shown in Figure 10.

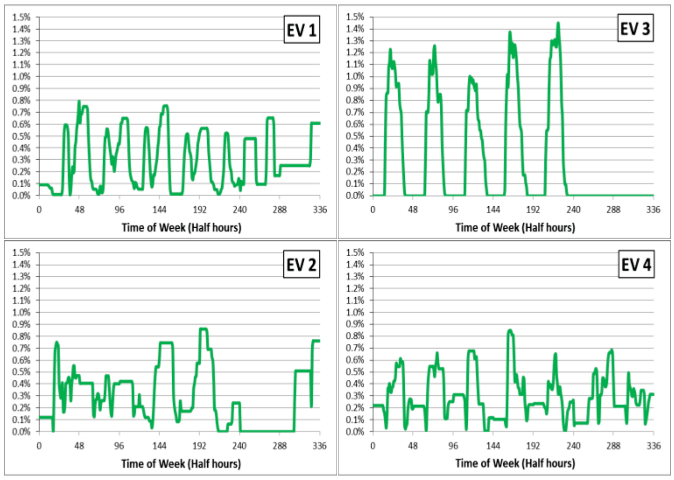

Figure 10: Charging Demand distribution for one week

Due to the small size of the sample, a generalisation was necessary in order to build a larger EV fleet. For this reason, the distributions of the analysed features were used to create EVs with similar behaviour. In this scenario 2,130 EVs were created, and the total charging demand of this fleet was calculated for one year. This charging demand was used as input to the forecasting model. The attributes of the Train and Test Datasets are illustrated below:

1. *Previous Week Load*: the charging demand of the same day of previous week for each half hour.
2. *Week:* number of the week (1-53).
3. *Day:* number of the day (1-7) starting with Monday.
4. *Half Hour:* 1-48 half hour parts of each day.
5. *Number of New Connections:* the new EV plug-in connections for every half hour.

Once the dataset is properly formed, the last week was selected for the Test Dataset, and all the previous ones for the Train Dataset.

3.2.2 Results

Using again the same procedure as described in Figure 5, the forecasts for one week were produced by the four data mining methods.

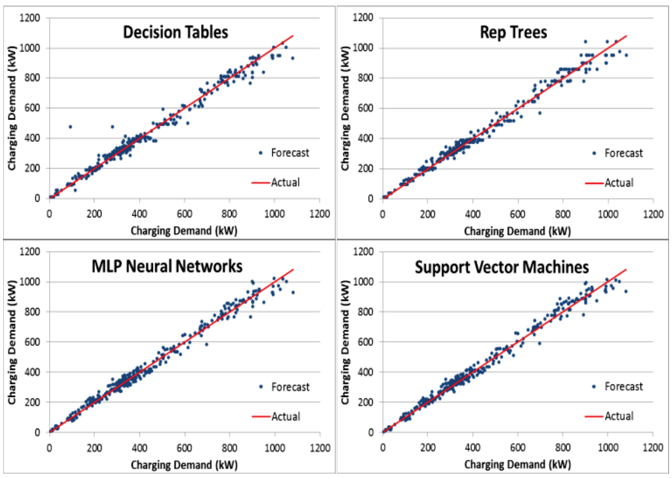

Figure 11: Correlation plots for each method

The correlation plots between the forecast and the actual values are represented in Figure 11, while the performance measures for each method are summarized in Table 4.

Method / Index	Decision Tables	Rep Trees	MLP ANN	SVM
MAPE (%)	7.975	6.850	6.975	6.74
RMSE	37.64	30.63	23.44	21.64
r-Correlation (%)	98.97	99.31	99.36	99.39
Training Time (s)	0.28	0.13	91.59	55.72
Testing Time (s)	0.09	0.08	0.11	1.06

Table4: Performance measures for each method

The above results showed that the MAPE increased. This is justified by the fact that the charging events on this case occurred in public charging stations. The fluctuation and randomness of the events are higher than in the residential charging stations, and so the relative errors increase. SVM provided again the most accurate forecast, but Rep Trees can reach almost the same accuracy needing zero time.

4 Conclusions

In this paper the use of data mining methods for forecasting the EV charging demand was studied. Two different realistic study cases where considered, and the performance of four different data mining methods was evaluated. In the first study case the day-ahead charging demand of 3,000 EVs was forecasted and compared to the actual data. The second study case considered a fleet of 2,130 EVs, and predicted the charging demand of a whole week on a half-hourly basis. The results showed that data mining methods can be used for forecasting the EV charging load, with increased accuracy especially when the configuration parameters of each method are carefully selected. However, more cases have to be studied, in order to clearly understand the key attributes that indicate the choice of one data mining method over another.

Acknowledgements

Information in the first study case is based upon work provided by The EV Project and supported by the U.S. Department of Energy under Award Number DE EE0002194 to ECOtality North America. The authors would also like to acknowledge "Schneider Electric" for the data of the second study case. This research was supported by the Engineering and Physical Sciences Research Council (EPSRC) which funded the project "Smart Management of Electric Vehicles".

References

[1] P. Papadopoulos, S. Skarvelis-Kazakos, I. Grau, L. M. Cipcigan, and N. Jenkins, "Predicting Electric Vehicle impacts on residential distribution networks with Distributed Generation", IEEE Vehicle Power and Propulsion Conference, Lille, France, 2010.

[2] "Department for Transport." [Online]. Available: http://www.dft.gov.uk/. [Accessed: 15-June-2013].

[3] Christopher Yang, Ryan McCarthy, "Electricity Grid: Impacts of Plug-In EV Charging", [Online]. Available: http://publications.its.ucdavis.edu/publication_detail.php?id=1290 [Accessed: 15-June-2013].

[4] Papadopoulos, O. Akizu, L. M. Cipcigan, N. Jenkins, E. Zabala, "Electricity Demand with Electric Cars: Comparing GB and Spain", Proc. IMechE Vol. 225 Part A: J. Power and Energy, pp.551-566, 2011.

[5] Cipcigan LM, Wells PN, Nieuwenhuis PA, Davies HC, Whitmarsh LE, "Electricity as a transportation fuel: Bridging the gaps in the electric vehicle value chain", EVVC-2012 European Electric Vehicle Congress, 2012.

[6] L. Pieltain Fernandez, T. Gomez San Roman, R. Cossent, C. Mateo Domingo, P. Frias, "Assessment of the Impact of Plug-in Electric Vehicles on Distribution Networks", IEEE Trans. Power Systems, vol. 26, pp. 206-213,Feb. 2011.

[7] A. G. Anastasiadis, E. Voreadi, and N. D. Hatziargyriou, "Probabilistic Load Flow Methods with High Integration of Renewable Energy Sources and Electric Vehicles – Case Study of Greece", IEEE PowerTech Conference, Trondheim, June 2011.

[8] Q. Gong, S. Midlam-Mohler, V. Marano, G. Rizzoni, " Study of PEV Charging on Residential Distribution Transformer Life", IEEE Trans. Smart Grid, vol. 3, issue 1, March 2012

[9] A.S. Masoum, S. Dei, P. S. Moses, A.A. Siada, "Voltage Profile and THD Distortion of Residential Network with High Penetration of Plugin Electrical Vehicles", IEEE/PES Innovative Smart Grid Technologies Conference Europe (ISGT Europe), Gothenburg, Sweden, 2010.

[10] K. Clement-Nyns, E. Haesen, J. Driesen, "The impact of charging plugin hybrid electric vehicles on a residential distribution grid", IEEE Trans. Power Systems, vol 25, pp. 371-380, Feb. 2010.

[11] E. Karfopoulos, C Marmaras and N.Hatziargyriou, "Charging Control Model for EV Supplier Aggregator", IEEE/PES Innovative Smart Grid Technologies Conference Europe (ISGT Europe), Berlin, 2012.

[12] Jiawei Han, MichelineKamber, "Data Mining: Concepts and Techniques", 2nd edition, Elsevier-Kaufmann, Chapter 1, 2006.

[13] A. Lojowska, D. Kurowicka, G. Papaefthymiou, and L. van der Sluis, "Stochastic Modeling of Power Demand Due to EVs Using Copula", IEEE Transactions on Power Systems, vol. 27, no. 4, pp. 1960–1968, Nov. 2012.

[14] F. Xie, M. Huang, W. Zhang, and J. Li, "Research on electric vehicle charging station load forecasting", International Conference on Advanced Power System Automation and Protection (APAP), pp. 2055–2060, Beijing, China, 2011.

[15] N. Jewell, M. Turner, J. Naber, and M. McIntyre, "Analysis of forecasting algorithms for minimization of electric demand costs for electric vehicle charging in commercial and industrial environments", IEEE Transportation Electrification Conference and Expo (ITEC), Michigan, USA, 2012.

[16] Joe H. Chow, Felix F. Wu, James Momoh, "Applied Mathematics for Restructured Electric Power Systems", Springer, Chapter 12, 2006.

[17] Ron Kohavi, "The Power of Decision Tables", European Conference on Machine Learning, 1995.

[18] U. W. Pooch, "Translation of Decision Tables", Computing Surveys, Vol. 6, No.2, June 1974.

[19] T. Elomaa and M. Kaariainen,"An analysis of reduced error pruning" Journal of Artificial Intelligence Research, 15:163–187, 2001.

[20] G. K. Jha, "Artificial Neural Networks", [Online]. Available: http://www.iasri.res.in/ebook/eb_smar/e-book_pdf%20files/manual%20iv/3-ann.pdf, [Accessed: 15-June-2013].

[21] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers", Proceedings of the fifth annual workshop on Computational learning theory (COLT '92), pp. 144–152, Pennsylvania, USA, 1992.

[22] M. Hall, E. Frank, G. Holmes, B.Pfahringer, P.Reutemann, I. H. Witten,"The WEKA Data Mining Software: An Update", SIGKDD Explorations, Volume 11, Issue 1. 2009

[23] "The EV Project", [Online]. Available: http://www.theevproject.com/downloads/documents/Q4%202012%20EVP%20Report.pdf [Accessed: 15-June-2013].

[24] S. Tzafestas and E. Tzafestas, "Computational Intelligence Techniques for Short-Term Electric Load Forecasting", Journal of Intelligent and Robotic Systems, vol.31, pp. 7-68, 2001.

[25] A. Desouky, "Accurate Fast Weather Dependent Load Forecast For Optimal Generation Scheduling In Real Time Application", Ph.D. dissertation, Bath University, UK, 2010.