ELSEVIER

AppliedEnergy

A data-driven approach for characterising the charging demand of electric vehicles: A UK case study

CrossMark

Erotokritos Xydas [a,*], Charalampos Marmaras [a], Liana M. Cipcigan [a], Nick Jenkins [a], Steve Carroll [b], Myles Barker [b]

[a] *Cardiff University, School of Engineering, The Queen's Buildings, The Parade, CF24 3AA Cardiff, Wales, UK*
[b] *Cenex, Innovation Centre, Loughborough University Science & Enterprise Parks, Oakwood Drive, LE11 3QF Loughborough, UK*

HIGHLIGHTS

- 21,918 charging events from 255 different charging stations in UK were analysed.
- A data pre-processing methodology for dealing with EVs charging data was presented.
- A data mining model was developed to analyse the EVs charging data.
- A fuzzy logic decision model was developed to characterise the EVs charging demand.

ABSTRACT

As the number of electric vehicles increases, the impact of their charging on distribution networks is being investigated using different load profiles. Due to the lack of real charging data, the majority of these load impact studies are making assumptions for the electric vehicle charging demand profiles. In this paper a two-step modelling framework was developed to extract the useful information hidden in real EVs charging event data. Real EVs charging demand data were obtained from Plugged-in Midlands (PiM) project, one of the eight 'Plugged-in Places' projects supported by the UK Office for Low Emission Vehicles (OLEV). A data mining model was developed to investigate the characteristics of electric vehicle charging demand in a geographical area. A Fuzzy-Based model aggregates these characteristics and estimates the potential relative risk level of EVs charging demand among different geographical areas independently to their actual corresponding distribution networks. A case study with real charging and weather data from three counties in UK is presented to demonstrate the modelling framework.

© 2015 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

Electric Vehicles (EVs) offer reduced transportation related emissions, reduce the energy cost of driving and in some cases eliminate the use of fossil fuels. The total electricity demand is expected to grow as the number of EVs increases [1]. The impact of EVs charging on distribution networks has been investigated in the literature. The majority of these studies use synthetic data to assess the impact of the EVs charging load due to limited access to real EVs charging data. In [2–19] data from travel surveys are used to create EVs charging load profiles, assuming that EVs are travelling like conventional internal combustion engine vehicles.

Although EVs adoption is at an early stage, some utilities and aggregators are already collecting information from charging stations. A limited number of EVs pilots exist around the world, allowing some preliminary studies on charging demand profiles. In [20], statistical analysis of 4933 charge events in the Victorian EVs Trial in Australia was performed. Statistical models for charge duration, daily charge frequency, energy consumed, start time of charge event, and time to next charge event were estimated to express the uncertainty of usage patterns due to different user behaviours. Data from the Western Australian Electric Vehicle Trial (2010–2012) were analysed in [21,22], investigating the drivers' recharging behaviours and patterns. In [23], 7704 electric vehicle recharging event data from the SwitchEV trials in the north east of England were used to analyse the recharging patterns of 65 EVs. The results showed that minimal recharging occurred during off peak times. In [24] data from the same project were combined

* Corresponding author.
   E-mail address: xydase@gmail.com (E. Xydas).

with low voltage smart meter data from Customer Led Network Revolution (CLNR) project and the impact of the combined demand profile was assessed on three different distribution networks. The results showed that the spatial and temporal diversity of EVs charging demand reduce its impact on those distribution networks. Finally, data from over 580,000 charging sessions and from 2000 non-residential electric vehicle supply equipment's (EVSE) located in Northern California were analysed in [25]. The scope of this analysis was to investigate the potential benefits of smart charging utilising the extracted information regarding the actual trips and customer characteristics.

Monitoring the charging events will inevitably create large volumes of data. These data require effective data mining methods for their analysis in order to extract useful information. In [26–28] various data mining techniques were utilized to address challenges in the energy sector, such as load forecasting and profiling. In [29–31] data mining modelling frameworks were applied to electricity consumption data to support the characterisation of end-user demand profiles.

In this paper, a framework was developed to characterize the EVs charging demand of a geographical area. The technical contributions of this paper are summarised below:

(i) Real EVs charging data from UK were acquired and analysed. The diverse data were organised and classified into attributes. To the authors' best of knowledge, this is the first time that real EVs charging data are presented using this level of detail.

(ii) A comprehensive data cleaning and formatting methodology is presented, developed specifically for dealing with EVs charging data.

(iii) A data mining model was developed to extract the useful information. Three key characteristics of EVs charging demand in a geographical area were investigated using the proposed methodology, namely shape of the typical daily profile, predictability with respect to weather and trend. Clustering, correlation and regression analysis were performed to study each characteristic, using factors to quantify them. Analysing these characteristics resulted in assessing the potential risks and uncertainties which affect the mid-term normal operation of the corresponding distribution network.

(iv) A fuzzy logic decision model was developed that aggregates the three factors into one "risk level" index. The "risk level" index was defined in order to characterize the EVs charging demand, reflecting its potential impact on the energy demand in a geographical area. Areas with high "risk level" values imply a potential risk for the mid-term normal operation of the distribution networks and such analysis could be important for the distribution network operator (DNO). No similar research work that quantifies the mid-term relative risk of the EVs charging demand among different geographical areas independently to their actual corresponding distribution networks was done so far.

(v) Furthermore, this paper fills a gap in the literature related to handling real EVs charging data, by proposing a complete data analysis methodology.

The rest of the paper is organized as follows: Section 2 describes the real EVs charging data analysed. In Section 3 the proposed methodology to characterize the EVs charging demand is illustrated. A case study is presented in Section 4, applying the model on real EVs charging events from UK to study the charging demand characteristics, and assess their potential impact. Finally, conclusions are drawn in Section 5.

2. Data description

EVs charging demand data were obtained from the Plugged-in Midlands (PiM) project (http://www.pluggedinmidlands.co.uk/). The Plugged-in Midlands project, managed by Cenex, is one of the eight 'Plugged-in Places' projects supported by OLEV, the Office for Low Emission Vehicles in the UK. Two datasets were provided by Cenex, with information regarding the charging events and charging stations respectively. The charging events dataset consists of 21,918 charging events from 255 different charging stations and 587 unique EVs drivers. The charging event dataset includes information about the connection/disconnection times and the energy of each charging event for the period of 2012–2013 with event-occurrence granularity. The charging station dataset contains time-independent information regarding the location and technical specifications of all charging points (e.g. the charging power rate). The contents of the two datasets are listed in Tables 1 and 2.

An additional dataset was acquired from the UK Met Office, with information regarding the weather in the Midlands, the geographical area under study. This dataset includes the values of various weather information (e.g. air temperature) with daily granularity for the period of 2012–2013. The weather attributes are listed in Table 3.

3. Methodology

The characterisation framework consists of three models: (i) *Data Pre-processing Model,* (ii) *Data Mining Model* and (iii) *Fuzzy Based Characterisation Model.* The *Data Pre-processing Model* provides data merging, cleaning and formatting to prepare the data

Table 1
Charging event data.

Attribute name	Attribute description
Connection time	Start time of charging event in dd/mm/yyyy hh:mm format
Disconnection time	End time of charging event in dd/mm/yyyy hh:mm format
Energy drawn	Energy demand of charging event in kW h
User	Unique ID for every EVs, e.g. EV1, EV2 etc.
Charging station	Unique ID for every charging station

Table 2
Charging station data.

Attribute name	Attribute description
Charging station	Unique ID for every charging station
Latitude	Latitude of charging station's location
Longitude	Longitude of charging station's location
Road	The road name of charging station's location
Post code	The post code of charging station's location
County	The county name of charging station's location
Location category	e.g. Private Parking, Public Parking etc.
Location subcategory	e.g. Public Car Park, Public On-street etc.
Ownership	e.g. Dealership, Hotel, Train Station
Host	Name of the charging station host
NCR	Whether or not the charging station is registered on the National Charging Registry (NCR) of UK
Manufacturer	The charging station manufacturer
Supplier	The operator of charging station
Charger type	Power rate of charging station in kW
Connector1	Socket Pin Type e.g. 3 Pin, 5 Pin etc.
Connector2	If exists, the second Socket Pin Type
Mounting type	e.g. Ground, Wall, Wall (tethered)

Table 3
Weather data.

Attribute name	Attribute description
Max air temperature	Daily maximum air temperature (°C)
Min air temperature	Daily minimum air temperature (°C)
Mean air temperature	Daily average air temperature (°C)
Mean wind speed	Daily average wind speed (knots)
Max gust	Daily maximum wind speed (knots)
Rainfall	Daily precipitation (mm)
Daily global radiation	Daily amount of solar energy falling on a horizontal surface (kJ/m²)
Daily sunny hours	Daily sunshine duration (h)

for the Data Mining model. The *Data Mining Model* consists of three modules namely Clustering Module, Correlation Module and Regression Module. These modules were used to investigate the shape of the typical daily profile, the predictability with respect to weather and the trend of EVs charging demand respectively. The *Fuzzy Based Characterisation Model* aggregates the outputs of the Data Mining model into a "risk level" index of EVs charging demand in a geographical area using fuzzy logic. The characterisation framework is illustrated with Fig. 1.

3.1. Data pre-processing model

Data of the Connection Time, Disconnection Time, Energy Drawn, Charging Station ID, Charger Type and County were selected and merged into one dataset (EV dataset). The EV dataset and the weather dataset were cleaned, removing missing and incorrect values. In the EV dataset, charging events with zero/negative energy were removed from the dataset. Charging events with average charging power higher than the nominal charger rate were corrected by calculating the actual charging duration using the nominal charger power rate. This consideration is based on the assumption that some EVs may be connected (parked) in a charging station but they are not charging. Therefore, the duration of EVs being connected to a charging station can be different to their actual charging duration. Duplicate data entries were also discovered and removed from both datasets.

Data regarding a charging event is recorded from the charging station and then forwarded to one or more data collection centres. This process involves a number of components and communication links increasing the risk of a potential failure in this chain.
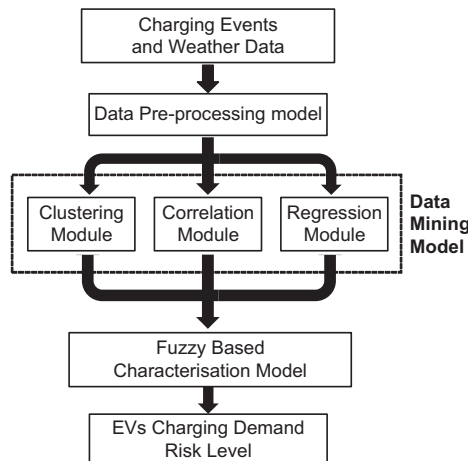
Corrupted or missing data are not a rare phenomenon in such complex communication networks. However, a careful analysis at this stage is also beneficial to find the location or the station's ID from where the corrupted data are recorded, an indication of an abnormal operation.

The next stage of the Data Pre-processing model is the Formation stage. The EV dataset was formatted using a Matlab script into three time series; an hourly power time series, a daily peak power time series and a monthly energy time series. The hourly power time series was transformed into daily vectors (each of 24 values) and forwarded to the Clustering Module, whereas the monthly energy time series was forwarded to the Regression Module. All the data attributes of the Weather dataset were formatted into daily time series and merged with the daily peak power time series. The resulting (combined) time series was forwarded to the Correlation Module. The data pre-processing procedure is presented in Fig. 2.

3.2. Data mining model

The Data Mining Model consists of a Clustering Module, Correlation Module and Regression Module. These modules were used to investigate the shape of the typical daily EVs charging demand profile, the predictability with respect to weather and the trend of EVs charging demand respectively.

3.2.1. Clustering module
The clustering module creates typical daily EVs charging demand profiles of a geographical area, according to the load demand of the corresponding charging stations. These profiles are related to the aggregated daily pattern of the EVs charging demand of a specific geographical area.

The k-means clustering method described in [32,33], was used in this module. Initially, this algorithm selects k random daily vectors (Input from Data Pre-processing Model) as the initial cluster centroids and calculates the distance from each daily vector to the cluster centroids. Each daily vector is assigned to a cluster/group
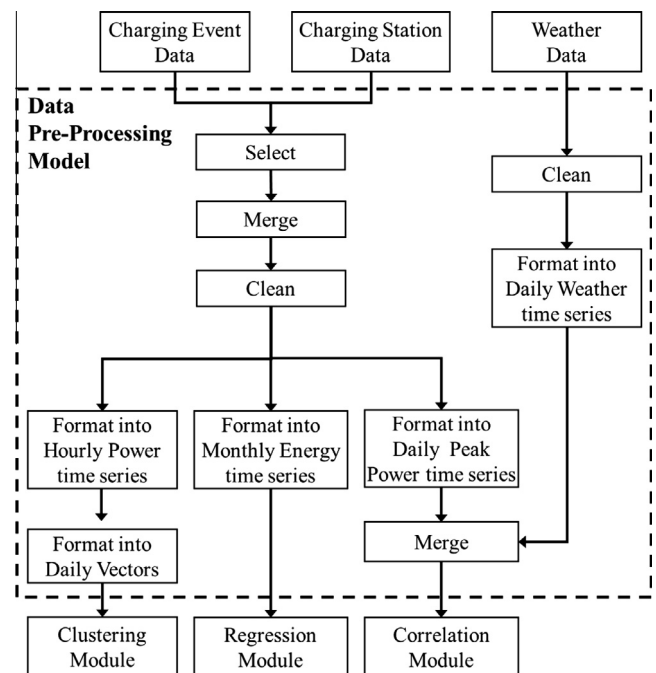

Fig. 1. Characterisation framework.

Fig. 2. Data pre-processing model.

according to its distance with the nearest cluster centroid. Then, the new cluster centroids are obtained from the average of the daily vectors for the corresponding cluster. This process is repeated until the distances between the daily vectors and the corresponding cluster centroids are minimized. This is explained mathematically by Eq. (1):

$$\min_{\mathbf{c}} \sum_{i=1}^{k} \sum_{\mathbf{x} \in c_i} \|\mathbf{x} - \mu_i\|^2 \qquad (1)$$

where c_i is the set of daily vectors that belong to ith cluster, $\mathbf{x}$ expresses the corresponding daily vector in c_i and μ_i is the position of the ith cluster centroid.

The method requires the number k of clusters to be defined a priori. The Davies–Bouldin evaluation criterion was used to calculate the number k of clusters [34,35]. This criterion is based on a ratio of within-cluster and between-cluster distances and is defined by Eq. (2):

$$DB = \frac{1}{k} \sum_{i=1}^{k} \max_{j \neq i} \left(\frac{\bar{d}_i + \bar{d}_j}{d_{ij}} \right) \qquad (2)$$

where $\bar{d}_i$ is the average distance between each point in ith cluster and the centroid of ith cluster. $\bar{d}_j$ is the average distance between each point in ith cluster and the centroid of jth cluster. d_{ij} is the distance between the centroids of ith and jth clusters. The maximum value of this ratio represents the worst-case within-to-between cluster ratio for ith cluster. The "best" clustering solution has the smallest Davies–Bouldin index value. Therefore, an additional step exists to evaluate the centroid selection for our dataset. A range of 1–20 clusters was considered, where 20 was found to be a reasonable maximum value [36], and the best number of clusters within this interval was calculated using an iterative process. By applying the k-means clustering method to the dataset, the k cluster centroids $\mathbf{c}_i$ are obtained, along with the number of vectors w_i assigned to each cluster. The followed steps of the Clustering Module are presented in Fig. 3.

The most representative cluster centroid (highest value of w_i) was used to create the typical daily EVs charging demand profile of an area. Having the daily EVs charging demand profile of an area, an index λ was defined to express the proportion of EVs charging demand during peak hours (17:00 – 20:00) [37]. The index λ was calculated using Eq. (3):

$$\lambda = \frac{E_{peak}}{E_{total}} \cdot 100\% \qquad (3)$$

where E_{peak} is the charging load during the peak hours and E_{total} is the total daily charging load.

3.2.2. Correlation module

According to [38], weather affects road traffic congestion and the driving behaviour of car owners. In [39–41], the factors which affect the fuel consumption of EVs were analysed. Cold weather decreases the efficiency of the batteries performance. Additionally, heating the interior of EVs drains significantly the battery. In [42], the impact of cold ambient temperatures on running fuel use was investigated. Considering EVs on the roads, the weather will also affect their energy consumption and thus their charging demand. Identifying hidden strong relationships between weather attributes and load demand improves the forecasting accuracy of a prediction model [43].

The Pearson's Correlation Coefficient (r) was used in this module to measure the correlation between the weather attribute values and the daily peak power of EVs charging demand in a geographical area. The maximum absolute correlation coefficient value of all peak power-weather pairs identifies the most influential weather attribute.

3.2.3. Regression module

The scope of this module is to investigate the monthly change of the EVs charging demand. A Growth Ratio (GR) index was defined as the ratio between the growth rate of EVs charging demand and the average monthly EVs charging demand. Linear regression analysis was applied on the EVs charging demand time series, in order to calculate the mathematical formula describing the relationship between monthly EVs charging demand (Y in kW h) and time (X in months). The formula is described with Eq. (4):

$$Y = \beta_0 + \beta_1 X + \varepsilon, \qquad (4)$$

where β_0 and β_1 are the constant regression coefficients and ε is the random disturbance (error).

The slope β_1 expresses the monthly growth rate of EVs charging demand (in kW h/month). The constant regression coefficients were calculated using the Least Squares Method described in [44]. Having β_1, the GR index is calculated with Eq. (5).

$$GR = \frac{\beta_1}{\overline{E}_{month}} \cdot 100\%, \qquad (5)$$

where $\overline{E}_{month}$ is the average monthly EVs charging demand.

3.3. Fuzzy based characterisation model

The goal of this model was to characterise the EVs charging demand of a geographical area according to the information about the shape of the typical daily profile (λ index), the predictability with respect to weather (r) and the trend of EVs charging demand (GR index). To this end, a "risk level" index was defined to aggregate the potential underlying risks from these characteristics. A fuzzy-logic model was developed to capture the fuzziness of these risks and calculate the "risk level" index. Fuzzy Logic Models are useful for risk assessment purposes under such conditions [45]. The *Fuzzy Based Characterisation Model* is illustrated with Fig. 4.

The validity of the risk characterisation model is based on the following considerations/assumptions:

i. The magnitude and duration of the peak of the typical EVs charging demand profile (captured by λ index) are underlying risk factors for the distribution network, as they affect the transformer/circuit loading and the voltage profile.

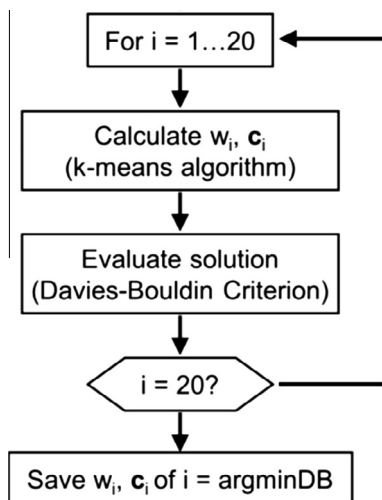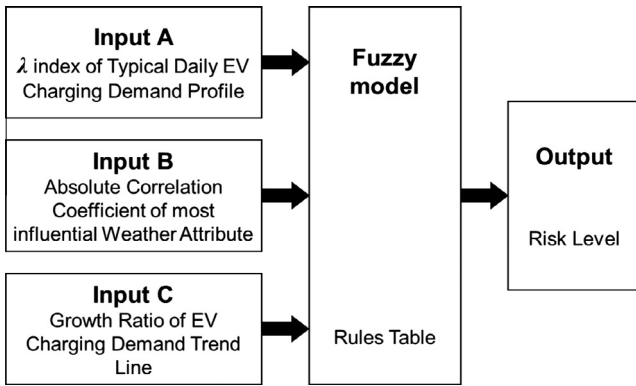**Fig. 3.** Clustering Module flowchart.

Fig. 4. Fuzzy Based Characterisation Model.

Fig. 6. Fuzzy membership function of Correlation Coefficient.

Fig. 7. Fuzzy membership function of EVs Demand Growth Ratio.

Fig. 8. Fuzzy membership function of "Risk Level" Factor.

ii. The change over time of EVs charging demand (described with GR index) affects the long term decision regarding the planning of the network reinforcement. The aggressiveness of EVs charging demand change over time in a geographical area is also a potential risk for the network's operation.

iii. The predictability of EVs charging demand with respect to weather in a geographical area (captured by r), affects the accuracy of a forecasting model. Decisions taken based on a forecast are subject to the forecasting accuracy, indicating a risk for the decision maker.

iv. Analysing the EVs charging demand characteristics in a geographical area results in assessing the risks and uncertainties which will affect the mid-term normal operation of the distribution network of the corresponding geographical area.

v. As an electric power network model was not used to analyse the related actual charging demand characteristics, this study quantifies only the relative risk between different geographical areas. The "risk level index" is not defined in absolute terms and thus it is used to classify relatively the level of these risks (due to EVs charging) among different geographical areas independently to their actual corresponding distribution networks.

The linguistic values used to express the input variables are Low (L), Medium (M) and High (H). Triangular membership functions are used to calculate the Degree-Of-Membership (DOM) for each of them, as shown in Figs. 5–7. In contrast to other kind of membership functions (e.g. Trapezoids), triangular membership functions are very sensitive to changes of the variables and thus this increase the accuracy.

The output is fuzzified into nine fuzzy regions represented by linguistic variables; very very high (VVH), very high (VH), high (*H*), medium high (MH), medium (M), medium low (ML), low (L), very low (VL) and very very low (VVL), as shown in Fig. 8. The rule table is given in Table 4.
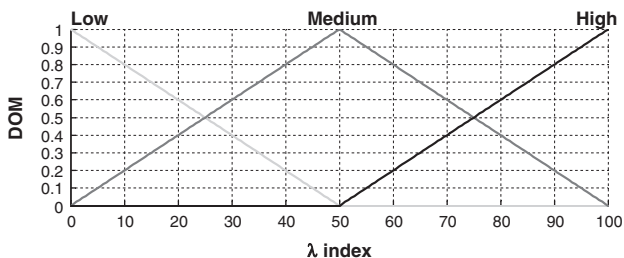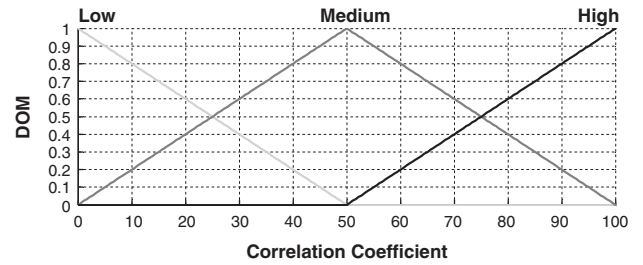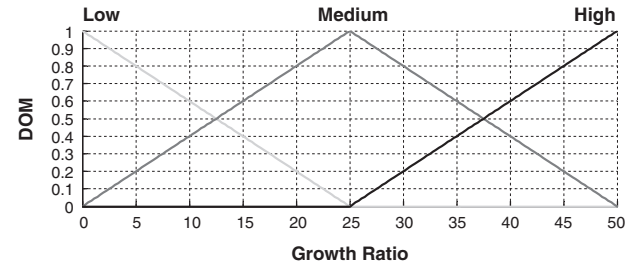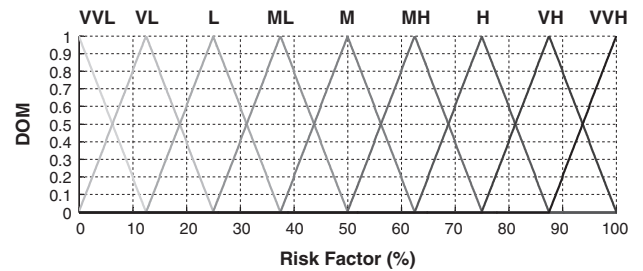
Table 4
Rule table.

r	λ	GR		
		L	M	H
H	L	VVL	VL	L
	M	VL	L	ML
	H	L	ML	M
M	L	L	ML	M
	M	ML	M	MH
	H	M	MH	H
L	L	M	MH	H
	M	MH	H	VH
	H	H	VH	VVH

The design of the rule table is based on the assumption that each of the input indicators affect equally the "risk level" index. According to the best of the authors' knowledge, there is no research work that quantifies the level of influence of the related indicators (λ index, r and GR) to the operation of an electricity distribution network. A further investigation is necessary to understand the relative impacts of these variables on the normal operation of an electricity distribution network, but this is out of the scope of this paper.

The Mamdani type inference was used (also known as the max–min inference method), which utilizes the minimum function for the implication of the rules. Defuzzification was performed using the centre of gravity (CoG) method [46–48]. This method finds

Fig. 5. Fuzzy membership function of λ index.

the centre of the area encompassed by all the rules, and thus the risk level index u is mathematically described by Eq. (6):

$$u = CoG = \frac{\int_{x_{min}}^{x_{max}} x \cdot g(x)dx}{\int_{x_{min}}^{x_{max}} g(x)dx} \qquad (6)$$

where x is the value of the "risk level" index, x_{min} and x_{max} represent the range of the "risk level" index and $g(x)$ is the degree of membership value at x.

4. Case study

The characterisation framework was applied on EVs charging data from three different geographical areas of the dataset. Charging events and weather data from the counties of Nottinghamshire, Leicestershire and West Midlands were analysed according the proposed modelling framework. Fig. 9 shows the locations of the charging stations for the corresponding geographical areas.

4.1. Typical EVs charging demand profiles

The k-means clustering algorithm was applied and the cluster centroids were obtained, along with their level of representation. Using the Davies–Bouldin criterion, the optimal number of clusters for Leicestershire was 5, for Nottinghamshire was 6 and for West Midlands was 3. The results are shown in Figs. 10–12.

The typical daily EVs charging demand profiles for each area are presented in Fig. 13. As seen from Fig. 13, the three typical EVs charging profiles differ in terms of peak magnitude, timing and duration. West Midlands shows the highest peak, however for a very short period (between 10:00 and 12:00), and no charging events during night. On the other hand, the typical EVs charging profiles of Nottinghamshire and Leicestershire have slightly lower peaks, but the charging activity takes place throughout the whole day. The EVs charging load during the peak hours, the total daily
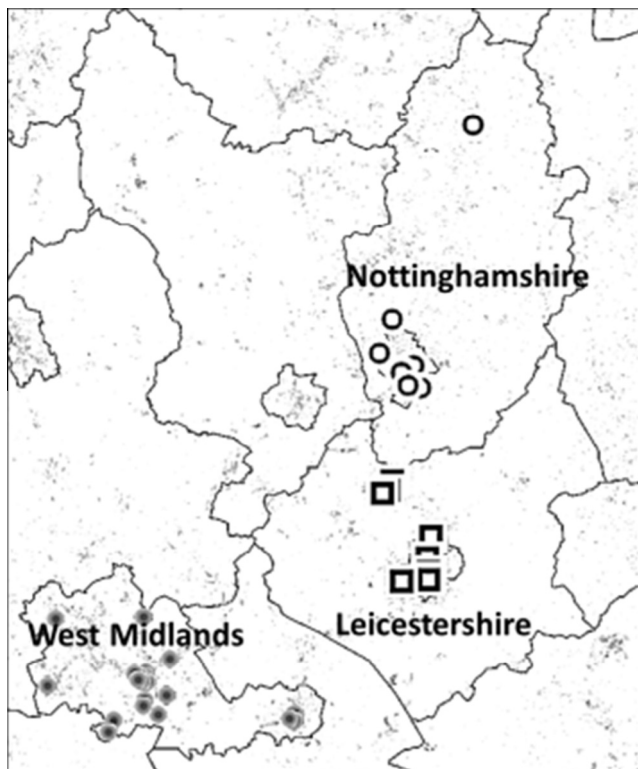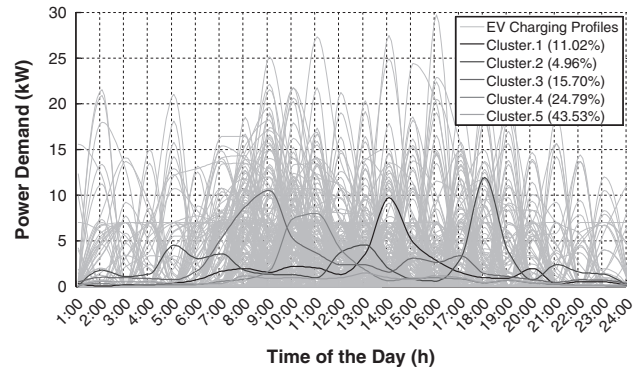

Fig. 10. Cluster centroids for Leicestershire.
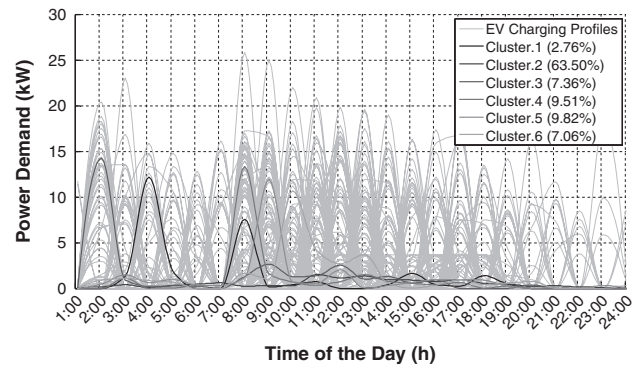
Fig. 11. Cluster centroids for Nottinghamshire.
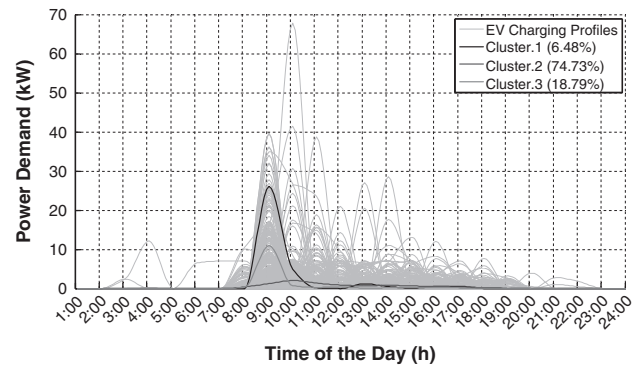
Fig. 12. Cluster Centroids for West Midlands.
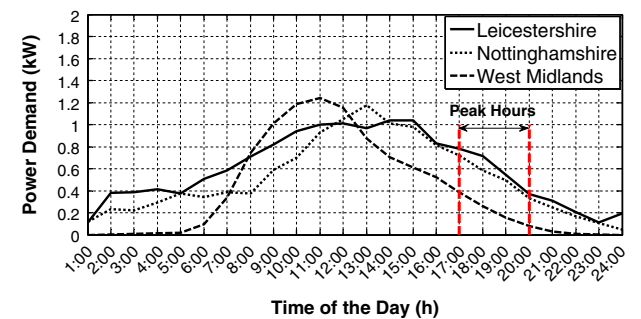

Fig. 9. Location from the analysed charging stations.

Fig. 13. Typical Profiles for all counties.

Table 5
Input values.

County	E_{peak}	E_{total}	λ (%)	Number of EVs	Number of charging events
Leicestershire	1.789	14.542	12.301	138	1944
Nottinghamshire	1.504	12.392	12.136	72	998
West Midlands	0.390	9.456	4.122	113	2013

charging load and their ratio λ are summarized in Table 5. The two last columns of Table 5 contain information about the total number of charging events and unique EVs drivers for the corresponding geographical areas.

As seen from Table 5, the proportion of the required energy during peak hours is relatively low for all counties. This is explained by the fact that the charging events are occurred in public charging stations. Public charging stations are expected to be used for recharging when EVs owners are at their work or when they do shopping or other activities. Considering the fact that the office hours are mostly between 09:00 and 17:00, the authors infer that most EVs owners return home after their work. Thus, this can be a possible justification why the energy requirements are low during peak times.

4.2. Influence of weather factors

Table 6 shows the absolute correlation coefficient (r) values between the weather attributes and the daily peak power of EVs charging demand. The most influential factor for all areas was temperature, with the Mean Air Temperature having the highest absolute correlation indices. Leicestershire's EVs charging demand shows a medium linear correlation, whereas in Nottinghamshire and West Midlands the EVs charging demand has a weaker relationship with weather.

As the above results show a dependency between EVs charging and Mean Air Temperature, it is useful to investigate the reasons for this relation. Although this investigation is out of the scope of this analysis, the authors provide their explanation about this dependency. In a northern country like UK the climate is considered cold and thus heating the interior of an electric vehicle will result in an increase of the energy requirements.

4.3. Trend of EVs charging demand

The linear regression module described in Section 2 was applied on the EVs charging demand time series of the three counties to calculate its growth rate. Figs. 14–16 present the daily EVs charging demand of each county for the period 2012–2013. Noticeable gaps exist in the data, especially for Leicestershire and West Midlands. The total monthly EVs charging demand is illustrated in Fig. 17, along with the corresponding trend line for each county.

Table 6
Correlation results.

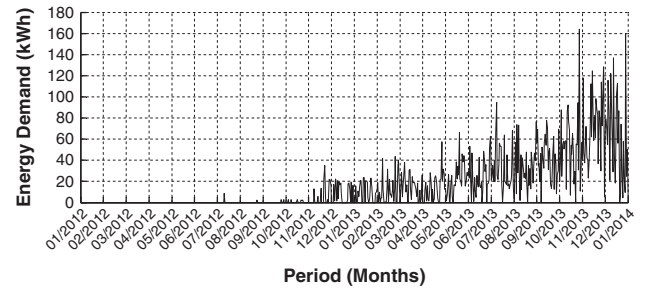| Weather attribute	County		
	Leicestershire (%)	Nottinghamshire (%)	West Midlands (%)
Max air temperature	26.18	14.66	15.58
Min air temperature	26.40	14.78	17.77
Mean air temperature	27.06	15.24	18.78
Mean wind speed	22.16	10.31	7.75
Max gust knots	12.83	5.57	10.88
Rainfall	7.54	1.80	0.20
Daily global radiation	11.00	1.91	5.63
Daily sunny hours	16.86	1.87	6.02

Fig. 14. Daily EVs charging demand for Leicestershire.
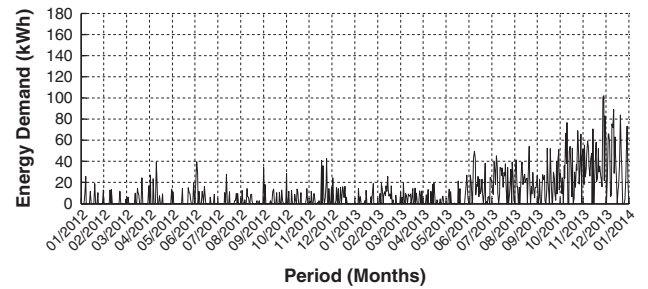

Fig. 15. Daily EVs charging demand for Nottinghamshire.
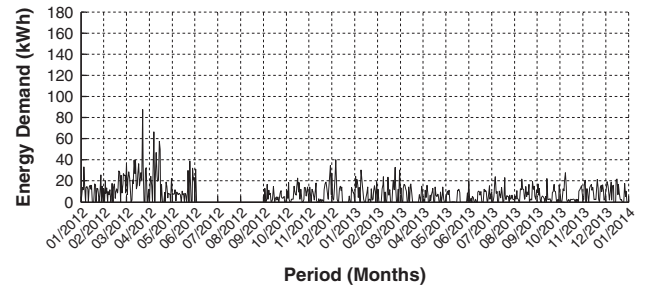

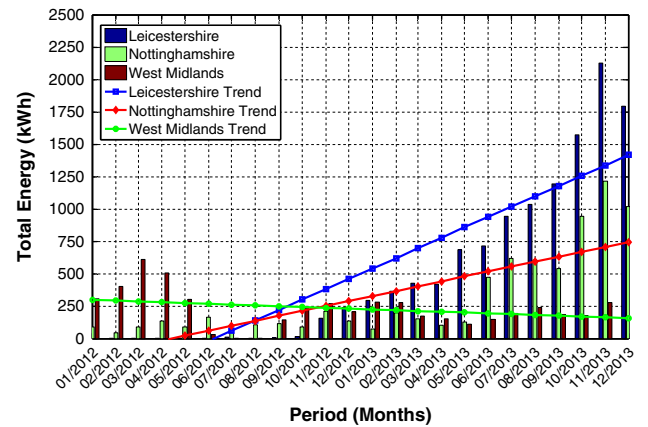**Fig. 16.** Daily EVs charging demand for West Midlands.

Fig. 17. Monthly EVs charging demand trend for all counties.

Using Eqs. (4), (5), the regression coefficients of the trend line were calculated along with the GR index for each county. The results are summarized in Table 7. As seen from the results, Leicestershire shows the highest EVs charging demand growth rate. On the

Table 7
Regression results.

County	β_1	β_0	GR (%)
Leicestershire	79.726	−496.75	15.95
Nottinghamshire	38.018	−166.94	12.33
West Midlands	−6.149	304.69	−2.69

Table 8
Input values.

County	Input A	Input B	Input C
Leicestershire	12.301	27.06	15.95
Nottinghamshire	12.136	15.24	12.33
West Midlands	4.122	18.78	−2.69

Table 9
Output values.

County	Risk index
Leicestershire	34.1
Nottinghamshire	34.8
West Midlands	23.5

contrary, the EVs charging demand in West Midlands reduces slightly over the two years period.

4.4. "Risk Level" calculation

Once the Data Mining process is completed, the *Fuzzy Based Characterisation Model* uses the outputs of the *Clustering, Correlation* and *Regression* modules to calculate the "risk level" index of EVs charging demand for each geographical area. Table 8 summarizes the input values for the characterisation model.

Input A is the λ index of each county's typical EVs charging demand profile, as calculated from the *Clustering module*. Input B is the absolute correlation coefficient (r) value of the EVs charging demand and Mean Air Temperature (the most influential weather factor), whereas Input C is the GR index of the EVs charging demand (monthly basis). The latter's membership function was assumed to accept values only in the range of [0%, 50%]; negative GR indices were assumed as 0% increase. The outputs of the *Fuzzy Based Characterisation Model* for the three counties are presented in Table 9.

As seen from Table 9, the EVs charging demand in West Midlands has the lowest value for "risk level" index. Looking at the corresponding input values, such a result is expected as the EVs charging demand has a descending trend (GR index) and low energy requirements during peak hours (λ index). Leicestershire and Nottinghamshire on the other hand are characterised with higher values of the risk level index by the model. Similar output values for these areas are not unexpected as Leicestershire has slightly higher growth ratio and energy requirements, however the EVs charging demand in Nottinghamshire is more unpredictable (lower correlation coefficient).

5. Conclusions

A characterisation framework for EVs charging demand was developed. The model utilizes data analysis methods to extract information hidden behind charging events in order to identify the characteristics of the EVs charging load. This information was then used by a fuzzy based characterisation model to estimate the underlying relative risks for the distribution networks among different geographical areas independently to their actual corresponding distribution networks. The framework was applied on a dataset of real charging events from three counties in UK and their "risk level" index was calculated.

The risk level index gives a spatial indication of the potential impact of the EVs charging demand on a distribution network in the nearby (mid-term) future. Areas with high "risk level" factor are candidates for further investigation. However, the interpretation of this index is highly influenced by the network characteristics. Other operational metrics (e.g. maximum load capacity) of the corresponding network should also be considered to plan possible network reinforcements. Charging strategies or other demand side management applications can be designed for an area according to its specific EVs charging load characteristics. For example, areas where the EVs charging demand is high during peak times, a valley filling strategy might be useful, whereas areas with random EVs charging events might need to invest on a different demand side management solution.

The universal design of this model makes it applicable from a county area to a neighbourhood area, as only minor changes are required for its application on different datasets. In addition, applying small modifications to the model, the analysis of additional EVs charging demand characteristics are easily supported.

References

[1] Element Energy Limited. Pathways to high penetration of electric vehicles: Final report for the Committee on Climate Change. Cambridge, 2013.

[2] Foley A, Tyther B, Calnan P, Gallachóir BÓ. Impacts of Electric Vehicle charging under electricity market operations. Appl Energy 2013;101:93–102.

[3] Mu Y, Wu J, Jenkins N, Jia H, Wang C. A Spatial-Temporal model for grid impact analysis of plug-in electric vehicles. Appl Energy 2014;114:456–65.

[4] Kelly JC, MacDonald JS, Keoleian GA. Time-dependent plug-in hybrid electric vehicle charging based on national driving patterns and demographics. Appl Energy 2012;94:395–405.

[5] De Gennaro M, Paffumi E, Scholz H, Martini G. GIS-driven analysis of e-mobility in urban areas: an evaluation of the impact on the electric energy grid. Appl Energy 2014;124:94–116.

[6] Harris CB, Webber ME. An empirically-validated methodology to simulate electricity demand for electric vehicle charging. Appl Energy 2014;126:172–81.

[7] Salah F, Ilg JP, Flath CM, Basse H, van Dinther C. Impact of electric vehicles on distribution substations: a Swiss case study. Appl Energy 2015;137:88–96.

[8] Wang H, Zhang X, Ouyang M. Energy consumption of electric vehicles based on real-world driving patterns: a case study of Beijing. Appl Energy 2015.

[9] Steen D, Tuan LA, Carlson O, Bertling L. Assessment of electric vehicle charging scenarios based on demographical data. IEEE Trans Smart Grid 2012;3 (3):1457–68.

[10] Shafiee S, Fotuhi-Firuzabad M, Rastegar M. Investigating the impacts of plug-in hybrid electric vehicles on power distribution systems. IEEE Trans Smart Grid 2013;4(3):1351–60.

[11] Darabi Z, Ferdowsi M. Aggregated impact of plug-in hybrid electric vehicles on electricity demand profile. IEEE Trans Sustain Energy 2011;2(4):501–8.

[12] Rezaee S, Farjah E, Khorramdel B. Probabilistic analysis of plug-in electric vehicles impact on electrical grid through homes and parking lots. IEEE Trans Sustain Energy 2013;4(4):1024–33.

[13] Papadopoulos P, Skarvelis-Kazakos S, Grau I, Cipcigan LM, Jenkins N. Electric vehicles' impact on British distribution networks. IET Electr Syst Transp 2012;2(3):91.

[14] ElNozahy MS, Salama MMA. A comprehensive study of the impacts of PHEVs on residential distribution networks. IEEE Trans Sustain Energy 2014;5 (1):332–42.

[15] Veldman E, Verzijlbergh RA. Distribution grid impacts of smart electric vehicle charging from different perspectives. IEEE Trans Smart Grid 2015;6 (1):333–42.

[16] Valsera-Naranjo E, Martinez-Vicente D, Sumper A, Villafafila-Robles R, Sudria-Andreu A. Deterministic and probabilistic assessment of the impact of the electrical vehicles on the power grid. IEEE Power Energy Soc General Meeting 2011;2011:1–8.

[17] De Gennaro M, Paffumi E, Martini G, Scholz H. A pilot study to address the travel behaviour and the usability of electric vehicles in two Italian provinces. Case Stud Transp Policy 2014;2(3):116–41.

[18] Pasaoglu G, Zubaryeva A, Fiorello D, Thiel C. Analysis of European mobility surveys and their potential to support studies on the impact of electric vehicles on energy and infrastructure needs in Europe. Technol Forecast Soc Change 2014;87:41–50.

[19] Ashtari A, Bibeau E, Shahidinejad S, Molinski T. PEV charging profile prediction and analysis based on vehicle usage data. IEEE Trans Smart Grid 2012;3 (1):341–50.

[20] Khoo YB, Wang C-H, Paevere P, Higgins A. Statistical modeling of electric vehicle electricity consumption in the Victorian EV Trial, Australia. Transp Res Part D Transp Environ 2014;32:263–77.

[21] Speidel S, Bräunl T. Driving and charging patterns of electric vehicles for energy usage. Renew Sustain Energy Rev 2014;40:97–110.

[22] Azadfar E, Sreeram V, Harries D. The investigation of the major factors influencing plug-in electric vehicle driving patterns and charging behaviour. Renew Sustain Energy Rev 2015;42:1065–76.

[23] Robinson AP, Blythe PT, Bell MC, Hübner Y, Hill GA. Analysis of electric vehicle driver recharging demand profiles and subsequent impacts on the carbon content of electric vehicle trips. Energy Policy 2013;61:337–48.

[24] Neaimeh M, Wardle R, Jenkins AM, Yi J, Hill G, Lyons PF, et al. A probabilistic approach to combining smart meter and electric vehicle charging data to investigate distribution network impacts. Appl Energy 2015.

[25] Kara EC, Macdonald JS, Black D, Bérges M, Hug G, Kiliccote S. Estimating the benefits of electric vehicle smart charging at non-residential locations: a data-driven approach. Appl Energy 2015;155:515–25.

[26] Pitt BD. Applications of data mining techniques to electric load profiling. University of Manchester, 2000.

[27] Dutta S. Data mining and graph theory focused solutions to Smart Grid challenges. 01-Dec-2012.

[28] Selbas R, Sencan A, Kucuksille E. Data mining method for energy system applications. In: Knowledge-oriented applications in data mining, InTech, 2011, p. 147–66.

[29] Figueiredo V, Rodrigues F, Vale Z, Gouveia JB. An electric energy consumer characterization framework based on data mining techniques. IEEE Trans Power Syst 2005;20(2):596–602.

[30] Zhang T, Zhang G, Lu J, Feng X, Yang W. A new index and classification approach for load pattern analysis of large electricity customers. IEEE Trans Power Syst 2012;27(1):153–60.

[31] Ramos S, Vale Z. Data mining techniques application in power distribution utilities. In: 2008 IEEE/PES Transmission and Distribution Conference and Exposition, 2008. p. 1–8.

[32] Hartigan JA. Clustering Algorithms. 1975.

[33] Tou JT, González RC. Pattern recognition principles. Addison-Wesley Pub Co; 1974.

[34] Davies DL, Bouldin DW. A cluster separation measure. IEEE Trans Pattern Anal Mach Intell 1979;PAMI-1(2):224–7.

[35] Petrović S, A Comparison Between the Silhouette Index and the Davies–Bouldin Index in Labelling IDS Clusters. 2008.

[36] Chicco G, Napoli R, Piglione F. Comparisons among clustering techniques for electricity customer classification. IEEE Trans Power Syst 2006;21(2):933–40.

[37] The Commission for Energy Regulation. Electricity Smart Metering Customer Behaviour Trials (CBT) Findings Report. Dublin, 2011.

[38] Cools M, Moons E, Wets G. Assessing the impact of weather on traffic intensity. Weather Clim Soc 2010;2(1):60–8.

[39] Smart J, Davies J, Shirk M, Quinn C, Kurani K, et al. Electricity demand of PHEVs operated by private households and commercial fleets: effects of driving and charging behavior. EVS25, Shenzhen, China 2010.

[40] Carlson R, Shirk M, Geller B. Factors affecting the fuel consumption of plug-in hybrid electric vehicles. In: The 25th international battery, hybrid and fuel cell electric vehicle symposium & exposition. Shenzhen, China, 2010, p. 289–94.

[41] Younes Z, Boudet L, Suard F, Gerard M, Rioux R. Analysis of the main factors influencing the energy consumption of electric vehicles. In: 2013 international electric machines & drives conference, 2013. p. 247–53.

[42] U.S. Environmental Protection Agency Office of Transportation and Air Quality. Fuel economy labeling of motor vehicles: revisions to improve calculation of fuel economy estimates. 2006.

[43] Taylor JW, Buizza R. Neural network load forecasting with weather ensemble predictions. IEEE Trans Power Syst 2002;17(3):626–32.

[44] Stigler SM. Gauss and the invention of least squares. Ann Stat 1981;9 (3):465–74.

[45] Zeng J, An M, Smith NJ. Application of a fuzzy based decision making methodology to construction project risk assessment. Int J Proj Manage 2007;25(6):589–600.

[46] Cox E. Fuzzy fundamentals. IEEE Spectr 1992;29(10):58–61.

[47] Dubois D, Prade H, editors. Fundamentals of fuzzy sets. US: Springer; 2012.

[48] Li HH, Gupta MM, editors. Fuzzy logic and intelligent systems. Netherlands: Springer; 1995.