Deep Learning-Based Probabilistic Forecasting of Electric Vehicle Charging Load with a Novel Queuing Model

Xian Zhang, K. W. Chan, Hairong Li, Huaizhi Wang, Jing Qiu, Guibin Wang

***Abstract*—With the emerging electric vehicle (EV) and fast charging technologies, EV load forecasting has become a concern for planners and operators of EV charging stations (CSs). Due to the non-stationary feature of traffic flow (TF) and the erratic nature of charging procedures, EV charging load is difficult to accurately forecast. In this study, TF is firstly predicted using a deep learning-based convolutional neural network (CNN), and different forecast uncertainties are evaluated to formulate the TF prediction intervals (PIs). Then, EV arrival rates are calculated according to historical data and the proposed mixture model. Based on TF forecasting and arrival rate results, the EV charging process is studied to convert the TF to the charging load using a novel probabilistic queuing model that takes into consideration of charging service limitations and driver behaviors. The proposed models are assessed using actual TF data, and the results show that the uncertainties of EV charging load can be learned comprehensively, indicating significant potential for practical application.**

***Index Terms*—Electric vehicle; convolutional neural network; deep learning; load forecast; driver behavior; queuing model.**

I. Introduction

A. Motivation

DUE to the rapid advancement of battery technologies, electric vehicles (EVs) have become a promising alternative for solving the current energy and environmental crises. The forecasting of EV charging load is essential for the planning and operation of EV charging facilities. Errors in long-term EV charging load forecasts can cause significant problems in the decision-making process for the construction of charging facilities [1]. Short-term uncertainties in EV charging load forecasts also impact the effect of the coordinated dispatch of EVs and renewables [2]. In general, there are two major kinds of barriers in EV load forecasting. First, charging procedures are not well-considered in EV load forecasting approaches. Second, the performances of the current forecasting approaches are still not satisfactory. Recently, efforts have been made to overcome such shortcomings [3]-[22] in terms of accurate and reliable forecasting of EV charging load.

B. Related Works

EV load forecasting based on historical data has been a common approach in the existing literatures [3]-[4]. A plug-in electric vehicle (PEV) sales forecasting model was developed in [3] based on consumer preferences for EVs, which were extracted from historical data. The daily EV load curve was then predicted based on EV charging behaviors. EV loads were forecasted in [4] using the number of EVs connected to the power grid, which was estimated using historical data from the National Renewable Energy Laboratory and Idaho National Laboratory. Due to the mobility characteristics of EVs, driving behavior should be considered in the EV load forecasting process [3]-[7]. A queuing theory based EV charging-demand stochastic model was proposed in [4], and a more accurate forecast result was obtained using real-time data and a proper analysis of uncertainties. The number of EVs using charging facilities during different hours of the day, charging start times, travel distances, and charging durations were considered to study demand forecasting for battery-swap stations in [5]. The half-hourly rolling vehicle to grid (V2G) capacity was estimated using dynamic real-time EV scheduling based on an accurate EV load model that considered constraints on meeting the demand, while ensuring EV charging ability in [6]. Due to the stochasticity of EV charging behavior, the Monte Carlo simulation has been frequently used in EV load forecasting [7].

Queuing theory [8] is also an effective approach for the study of aggregated EV charging behaviors. This approach considers various uncertainties in the EV charging process. The V2G capacity for regulation was estimated using queuing theory in [9]. Charging demand has also been forecasted based on the queuing model [6], [10]-[11]. However, applications of queuing theory that properly consider EV mobility and aggregative characteristics remain rare. Traffic flow (TF) theory [12] is a promising approach for solving this limitation and has gained rapid acceptance in EV modeling in recent years. EV TF is characterized using the Wardrop user equilibrium TF theory for the robust and coordinated operation of distribution systems coupled with EV charging facilities [37-38]. The Nesterov user equilibrium TF theory was used to model the steady-state distribution of EVs in the planning of electrified transportation networks [39]. The EV charging load spatial and temporal dynamics of a highway charging station (CS) were studied in [13] using queuing theory and a fluid dynamic TF model. The arrival rate was estimated using the spatial and temporal model based on the origin-destination analysis of TF in [14], and the CS capacity was obtained using the forecasted arrival rate and proposed queuing model.

A large number of TF forecasting studies have been performed due to the fast developments in smart traffic technologies. Various neural networks (NNs) including randomized learning NN [42], shallow NN [43], deep learning based NN [44], and NN without deep architecture [45] have been used in the forecasting studies. Parametric technologies,

This work is supported by National Natural Science Foundation of China (51507103, 51707123). X. Zhang acknowledges the support of Hong Kong Polytechnic University Research Studentship for her Ph.D. study. H. Li, H. Wang, and G. Wang are with the College of Mechatronics and Control Engineering, Shenzhen University, Shenzhen 518060, China. (e-mail wanghz@szu.edu.cn). X. Zhang and K. W. Chan are with the Department of Electrical Engineering, The Hong Kong Polytechnic University, Hong Kong. J. Qiu is with the School of Electrical and Information Engineering, The University of Sydney, Camperdown, NSW, 2006, Australia. (*Corresponding Author*: Huaizhi Wang)

such as the autoregressive integrated moving average [4] and other time-series approaches [15], have been widely used for TF predictions. Nonparametric approaches, such as the K-nearest neighbors approach [16] and the nonparametric regression [17], have also been broadly adopted due to their ability to integrate the stochastic characteristics of TFs [18]. However, deep learning has a better ability than the previously described approaches in discovering the inherent features from lowest level to the highest lever, given their good performance in the extracting of structures within TF data [19]-[22]. When comparing with existing approaches, deep learning approaches, such as stacked autoencoders (SAE) [19]-[20] or deep belief networks [21]-[22], represent the inner features of TFs without any prior knowledge, and they have superior TF prediction performance.

C. Contributions

This paper is devoted to addressing complicated EV load forecasting problems using a comprehensive CNN-based ensemble approach. Contributions of this work to the field include the following aspects. First, TF is forecasted using a deep-learning based convolutional neural network (CNN) approach in which an ensemble approach that considers both model and data uncertainty is employed to effectively formulate the TF prediction intervals (PIs). The prediction is a hybrid of wavelet decomposition (WT), CNN, and PI construction. The complicated nonlinear features of TFs are represented more effectively, and a superior forecasting performance is obtained. Second, a mixture model-based approach is used to approximate the arrival rate of EVs according to historical data. Third, an advanced queuing model is formulated for the first time to predict the EV charging load in the CSs, which scientifically accounts for CS service limitations and the inherent stochasticity of EV driver behaviors. The application of CNN on EV charging demand forecasting is novel and effective, and its accuracy and reliability is essential to the operation and control of the FCS, power system, and traffic system, e.g., to help the FCS operator to determine the dispatchable EVs in advance to avoid higher costs or excessive risks, to assist the traffic operator to alleviate traffic congestion, and to help the decision making for the electricity market trading through optimized bidding when EVs participant in ancillary service market, and so on.

II. Traffic Flow Forecast with CNN and PIs

Due to the chaotic nature of TF, EV charging loads possess high volatility and variability. Therefore, a convolution neural network (CNN) based comprehensive approach is introduced in this paper to reduce the influence of uncertainties on charging demand accuracy. The raw TF data are first normalized and decomposed into couples of frequencies. CNNs are then designed for each frequency and trained to predict the behavior of each frequency. Consequently, wavelet reconstruction and anti-normalization are used to synthesize prediction frequencies to obtain a deterministic point forecast of TF. PIs are formulated according to model and data uncertainties. Finally, charging load forecasting is conducted based on the proposed data estimation and queuing model. The flowchart of the CNN implementation is shown in Fig. 1, and the details are elaborated below.

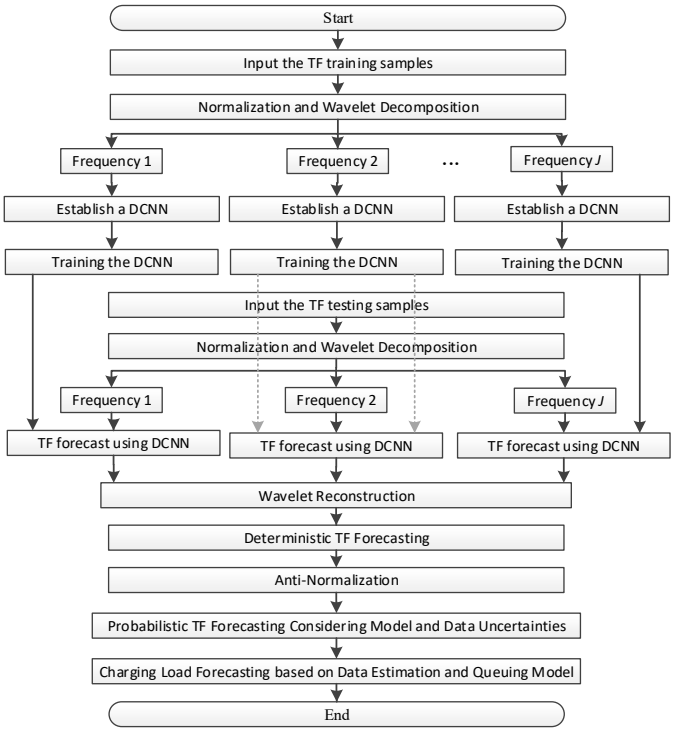

Fig. 1. The flowchart of the CNN implementation.

A. Convolutional Neural Network

CNN is a type of feed-forward artificial neural networks (ANN) that has gained popularity in face recognition and language processing applications. As shown in Fig. 2, a pair of convolution and pooling layers in succession is referred to as 1 *CNN* [23]. It can be seen in Fig. 2 that either the convolution or the pooling layer consists of several maps, and each map has several neurons that share the same neuron kernel (also called weight). The data processed by the CNN are converted to 2-dimensional maps and gives a compact representation of a large set of hidden features. This ability makes CNN learn in an effective way to extract nonlinear structures in TF.

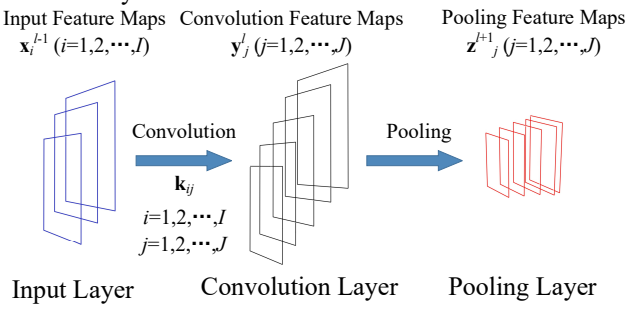

Fig. 2. Basic structure of a CNN.

(1) *Convolution Layer*:

The convolution layer handles small local receptive fields of input data in a sliding-map manner. At the current convolution layer l, the previous (l-1) layer's feature maps are convolved with learnable kernels and passed through the activation function to form the output feature map before being transmitted to the pooling layer. This process is described in general as follows:

$$\mathbf{y}_j^l = f(\sum_{i \in M_j} \mathbf{x}_i^{l-1} \otimes \mathbf{k}_{ij}^l + b_j^l) \quad (1)$$

where $\mathbf{y}_j^l$ is the output of the jth map in layer l; $\mathbf{x}_i^{l-1}$ is the input of the ith map in layer l-1; M_j represents a selection of input maps; $\otimes$ denotes the convolutional operation; and $\mathbf{k}_{ij}$ and b_j represent the weight and bias of the corresponding convolutional layer, respectively. Learning in an NN progresses by incrementally adjusting the biases and weights. The vector of weights and bias are defined as a *filter* and represents some feature of the input such as a kind of shape. The filter works in a sliding-window fashion on small local receptive fields of data. In this work, the sigmoid function, $S(x)=1/(1+e^{-x})$, is used as the activation function, $f(\cdot)$, and the weight, $\mathbf{k}_{ij}$, is trained with the well-known back-propagation algorithm.

(2) *Pooling Layer*:

A pooling layer generates several down-sampled versions of the input maps according to the following rules:

$$\mathbf{z}_j^{l+1} = f(\beta_j^{l+1} \mathrm{down}(\mathbf{y}_j^l) + c_j^{l+1}) \quad (2)$$

Each output map has its own multiplicative bias, β, and an additive bias, c. Down$(\cdot)$ represents a pooling function. Also, the average function is adopted in this work:

$$\mathbf{z}_{i,j,k} = \sum_{p,q} \boldsymbol{\alpha}_{p,q} \mathbf{y}_{i+p,j+q,k} \quad (3)$$

where $\mathbf{y}_{i,j,k}$ and $\mathbf{z}_{i,j,k}$ represent each element of $\mathbf{y}$ and $\mathbf{z}$, respectively, and $\boldsymbol{\alpha}_{p,q}$ is the average filter with size $p \times q$.

(3) *Advantages of CNN*

The CNN is consisting of the convolution layer and the pooling layer. The major benefits of CNN lie in its superior ability of feature representation [46], which are elaborated as follows: The convolution layer is quite different from the fully connected hidden layer of other learning approaches. Each convolutional unit represents some features of a local region of the input since it receives input only from a local area. The units of the convolution layer can be self-structured into a few feature maps, where all units in the same feature map share the same weights but receive inputs from different lower layer locations [23]. The unique feature of CNNs of many weights sharing the same filter improves performance significantly by reducing both the memory footprint and the number of parameters to be estimated [24]. This is because a single bias and a single vector of weights is used across all the receptive fields that share that filter, rather than each receptive field using its own bias and vector of weights [46]. This indicates that the CNN learns the filter in an easy and efficient way without too much pre-processing, and this independence from prior knowledge in feature representation is a major advantage of CNNs. Therefore, the architecture of the CNN involves more connections that weights and presents some degree of translation invariance [47], making it much easier to train. The CNN also has fewer parameters to be estimated compared with other NNs. Meanwhile, pooling layers intersperse convolution layers to reduce the computational burden and build up spatial and configural invariance. The pooling operation has 2 main advantages. First, it reduces the dimensionality of the convolutional layer output. Furthermore, the pooling summarizes the neighboring feature activations, leading to the robustness of the forecast by the translation of input data [25].

B. The Overall CNN Architecture

The overall CNN architecture is composed of multiple CNNs [23]. Due to the complications of the TF series, a novel modified overall CNN architecture is proposed in this study, and Fig. 3 shows a simple modified overall CNN architecture with wavelet decomposition, 2 convolution layers, and 1 final fully connected layer. First, original TF data series are usually nonlinear and non-stationary, and this will deteriorate traffic forecasting accuracy. Besides, the original TF data series are time-domain series which are not the best representation for the forecasting processing, and the information cannot be readily seen in the time-domain. Given that the most distinguished information is hidden in the frequency content of the series in many cases, decomposing the data series into several frequencies with better behavior of the data variance and outliers is helpful. In the beginning of the overall CNN architecture, a fast-discrete WT algorithm [26] is adopted for data decomposition to obtain a better forecast performance. WT could provide the time and frequency information simultaneously, hence giving a time-frequency representation of the series. This unique feature of WT means it has ability to determine all the frequencies in the data series and also knows when the frequencies are present. The WT, $W(m, n)$, of signal, $g(x)$, with respect to a mother wavelet, $\phi(x)$, is defined as:

$$W(m,n) = 2^{-(m/2)} \sum_{t=0}^{T-1} g(t) \phi\left[\left(t - n2^m\right) / 2^m\right] \quad (4)$$

where m and n denote 2 integer variables that determine the parameters of scaling and translation of ϕ; t is the discrete time index; and T is the length of the signal $g(t)$.

Next, the normalized and wavelet decomposed training samples are fed into the overall architecture containing consecutive multiples CNNs, the structure of which has been discussed in Section II. A. At the end of the overall architecture, several fully-connected hidden layers are added on top of the final CNN layer to combine the features across all maps before feeding to the output layer. A linear function is applied and the fully connected layer is stated as

$$\mathbf{y}^l = \mathrm{K}^l \mathbf{x}^{l-1} + \mathbf{b}^l \quad (5)$$

where K^l is the weight from layer l-1 to layer l; and $\mathbf{b}^l$ is the additive bias.

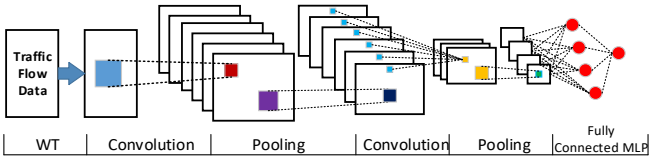

Fig. 3. The overall CNN architecture.

C. PIs Formulation and Performance Evaluation

(1) *An Overview of PI formulation*

The overall PIs formulation framework is demonstrated in Fig. 4. The estimated regression, $\hat{y}(\mathbf{x}_i)$, model uncertainty variance, σ_m^2, and data uncertainty variance, σ_d^2, are obtained using forecasters based on CNN training to formulate PIs and are elaborated as follows.

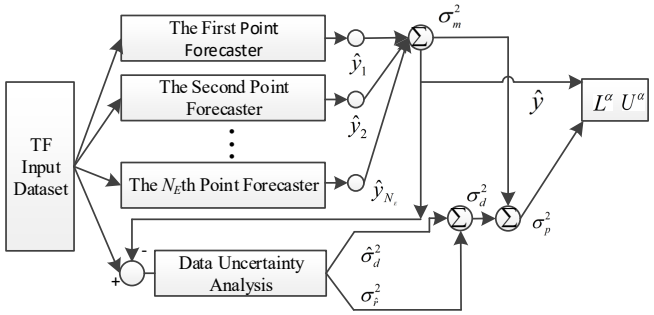

Fig. 4. An overview of PIs formulation.

(2) *Estimated regression*

The *i*th true forecast target, t_i, can be described as $t_i = y(\mathbf{x}_i) + \varepsilon(\mathbf{x}_i)$, where $\mathbf{x}_i$ is the vector of the input data; $y(\mathbf{x}_i)$ is the true regression mean; and $\varepsilon(\mathbf{x}_i)$ is the noise. According to the definition of mean value, the *estimated regression*, $\hat{y}(\mathbf{x}_i)$, of the trained CNN is the mean value of targets conditioned on $\mathbf{x}_i$, $E[t_i \mid \mathbf{x}_i]$ [40], which can be considered as an estimation of the true regression, $y(\mathbf{x}_i)$:

$$\hat{y}(\mathbf{x}_i) = E[t_i \mid \mathbf{x}_i] = (1/N_E)\sum_{q=1}^{N_E} \hat{y}_q(\mathbf{x}_i) \tag{6}$$

where q is the number of CNN models. Then values of $E[t_i \mid \mathbf{x}_i]$ can then be derived based on the trained N_E CNNs with the target t_i. An ensemble of CNN models with a larger number, N_E, can achieve a less biased estimate, $\hat{y}(\mathbf{x}_i)$, of the true forecast target, t_i.

(3) *Model uncertainty variance and data uncertainty variance*

There are two major types of uncertainties in CNN prediction: *model uncertainty* and *data uncertainty*. Model uncertainty usually originates from the erroneous specification of the CNN structure and parameters, local minimum of the training process, and finite training samples. Meanwhile, data uncertainty usually comes from the stochastic nature of the data noise, such as the noise caused by chaotic weather conditions.

The total forecast error, $\sigma_p(\mathbf{x}_i)$, is the difference between the forecast target, t_i, and the estimated regression, $\hat{y}(\mathbf{x}_i)$: $\sigma_p(\mathbf{x}_i) = t_i - \hat{y}(\mathbf{x}_i) = [y(\mathbf{x}_i) - \hat{y}(\mathbf{x}_i)] + \varepsilon(\mathbf{x}_i)$, and $[y(\mathbf{x}_i) - \hat{y}(\mathbf{x}_i)]$ is the error involved in the estimation of the true regression, $y(\mathbf{x}_i)$, which accounts for model uncertainty. According to the variance definition discussed by [27], the variance of the model uncertainty can be evaluated from the variance in the outputs of the trained N_E CNN models according to the following equation:

$$\sigma_m^2(\mathbf{x}_i) = (1/(N_E - 1))\sum_{q=1}^{N_E} (\hat{y}_q(\mathbf{x}_i) - \hat{y}(\mathbf{x}_i))^2 \tag{7}$$

By contrast, the error involved in the true forecast targets, t_i, accounts for the data uncertainty, which is difficult to estimate since there is only 1 observation of TF at each time point. Therefore, it is presumed in this work that the mean and variance of data uncertainty are also conditioned on $\mathbf{x}_i$. The variance of the measured target t_i conditioned on $\mathbf{x}_i$ are defined as the variance of data, $\sigma_d^2(t_i \mid \mathbf{x}_i)$, which is expressed as follows, according to the variance definition [27]:

$$\sigma_d^2(t_i \mid \mathbf{x}_i) = E[(t_i - E[t_i \mid \mathbf{x}_i])^2 \mid \mathbf{x}_i] \tag{8}$$

According to (6), $E[t_i \mid \mathbf{x}_i]$ can be replaced with $\hat{y}(\mathbf{x}_i)$, and the output of $\sigma_d^2(\mathbf{x}_i)$ can be expressed as:

$$\sigma_d^2(t_i \mid \mathbf{x}_i) = \hat{r}(\mathbf{x}_i) = E[(\hat{y}(\mathbf{x}_i) - t_i)^2 \mid \mathbf{x}_i] \tag{9}$$

Similar to the process for obtaining $E[t_i \mid \mathbf{x}_i]$, $E[(\hat{y}(\mathbf{x}_i) - t_i)^2 \mid \mathbf{x}_i]$ can be derived by training the CNN with the target $(\hat{y}(\mathbf{x}_i) - t_i)^2$, the regression of which is denoted as $\hat{r}(\mathbf{x}_i)$. $\sigma_d^2(\mathbf{x}_i)$ in (9) can then be derived.

To reduce the bias of data uncertainty, N_S CNNs with the regression, $\hat{r}(\mathbf{x}_i)$, are ensemble-trained, and the estimated noise variance is:

$$\hat{\sigma}_d^2(t_i \mid \mathbf{x}_i) = (1/N_S)\sum_{l=1}^{N_S} \hat{r}_l(\mathbf{x}_i) \tag{10}$$

The variance of $\hat{r}(\mathbf{x}_i)$ model uncertainty is:

$$\sigma_{\hat{r}}^2(\mathbf{x}_i) = (1/(N_S - 1))\sum_{l=1}^{N_S} (\hat{r}_l(\mathbf{x}_i) - \hat{r}(\mathbf{x}_i))^2 \tag{11}$$

Hence, the data uncertainty variance is the combination of (10) -(11):

$$\sigma_d^2(\mathbf{x}_i) = \hat{\sigma}_d^2(t_i \mid \mathbf{x}_i) + \sigma_{\hat{r}}^2(\mathbf{x}_i) \tag{12}$$

Based on the variance of model uncertainty, $\sigma_m^2(\mathbf{x}_i)$ in (7), and the variance of data uncertainty, $\sigma_d^2(\mathbf{x}_i)$ in (12), the variance of the total forecast errors is:

$$\sigma_p^2(\mathbf{x}_i) = \sigma_m^2(\mathbf{x}_i) + \sigma_d^2(\mathbf{x}_i) \tag{13}$$

(4) *PIs formulation based on forecast errors*

Therefore, the 100(1-α)% PI nominal confidence (PINC) of t_i, which is an interval denoted as $I^\alpha(\mathbf{x}_i)$=[$L^\alpha(\mathbf{x}_i)$, $U^\alpha(\mathbf{x}_i)$], can be formulated using the lower bound, $L^\alpha(\mathbf{x}_i)$, and upper bound, $U^\alpha(\mathbf{x}_i)$.

$$L^\alpha(\mathbf{x}_i) = \hat{y}(\mathbf{x}_i) - z_{1-\alpha/2}\sqrt{\sigma_p^2(\mathbf{x}_i)} \tag{14}$$

$$U^\alpha(\mathbf{x}_i) = \hat{y}(\mathbf{x}_i) + z_{1-\alpha/2}\sqrt{\sigma_p^2(\mathbf{x}_i)} \tag{15}$$

(14) and (15) are the mathematical expressions of confidence intervals, meaning that the target, t_i, is within the confidence interval [$L^\alpha(\mathbf{x}_i)$, $U^\alpha(\mathbf{x}_i)$] with a probability of 100(1-α)%. $z_{1-\alpha/2}$ is the critical value of the standard normal distribution and depends on the required PINC level 100(1-α)%.

Mean absolute error (MAE), mean absolute percentage error (MAPE), and root mean square error (RMSE) are employed to evaluate the deterministic forecast performance as follows

$$\mathrm{MAE} = (1/N)\sum_{i=1}^{N} |t_i - \hat{y}(\mathbf{x}_i)| \tag{16}$$

$$\mathrm{MAPE} = (1/N)\sum_{i=1}^{N} (|t_i - \hat{y}(\mathbf{x}_i)| / ((\sum_{i=1}^{N} t_i)/N)) \times 100\% \tag{17}$$

$$\mathrm{RMSE} = \sqrt{(1/N)\sum_{i=1}^{N} |t_i - \hat{y}(\mathbf{x}_i)|^2} \tag{18}$$

where N is the number of training samples. PI coverage probability (PICP) and average coverage error (ACE) are employed as reliability indices to evaluate how well the probabilistic forecast results match the observed values:

$$\mathrm{PICP} = (1/N)\sum_{i=1}^{N} r_i^\alpha \times 100\% \tag{19}$$

$$\mathrm{ACE} = \mathrm{PICP} - \mathrm{PINC} \tag{20}$$

where r_i^α is the PI coverage probability indicator

$$r_i^\alpha = \begin{cases} 1 & t_i \in I^\alpha(\mathbf{x}_i) \\ 0 & t_i \notin I^\alpha(\mathbf{x}_i) \end{cases} \tag{21}$$

The interval sharpness (IS) is used to measure the sharpness of the PI by encouraging narrower PIs

$$\mathrm{IS}=\frac{1}{N}\sum_{i=1}^{N}\begin{cases}-2\alpha\delta^{\alpha}(\mathbf{x}_i)-4(L^{\alpha}(\mathbf{x}_i)-t_i) & t_i<L^{\alpha}(\mathbf{x}_i)\\ -2\alpha\delta^{\alpha}(\mathbf{x}_i) & t_i\in I^{\alpha}(\mathbf{x}_i)\\ -2\alpha\delta^{\alpha}(\mathbf{x}_i)-4(t_i-U^{\alpha}(\mathbf{x}_i)) & t_i>U^{\alpha}(\mathbf{x}_i)\end{cases} \tag{22}$$

where $\delta^{\alpha}(\mathbf{x}_i)$ is the width of the PI and calculated as $U^{\alpha}(\mathbf{x}_i)-L^{\alpha}(\mathbf{x}_i)$.

Lastly, the pinball loss (PL) evaluates the accuracy of the probabilistic forecast results.

$$\mathrm{PL}=\begin{cases}(t_i-\hat{y}(\mathbf{x}_i))\alpha & t_i\geq\hat{y}(\mathbf{x}_i)\\ (\hat{y}(\mathbf{x}_i)-t_i)(1-\alpha) & \hat{y}(\mathbf{x}_i)>t_i\end{cases}$$

III. Queuing Based Probabilistic EV Load Forecast

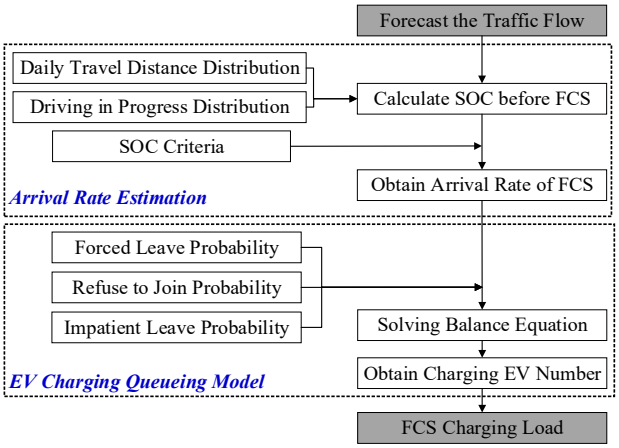

Fig. 5. Flowchart of the FCS charging load determination.

Although refueled by charging posts at home or in parking lots leads to less battery degradation, EV owners will also choose to charge using fast charging stations (FCS) when their vehicle urgently needs charging, or they are reluctant to wait for an extended period. Uncertainties are involved in this EV charging process. As shown in Fig. 5, a practical estimation approach for the arrival rate is proposed using the daily travel distance distribution and travel patterns. Then, an EV queuing model is established for EV load forecasting that considers various uncertainties and constraints.

A. EV Arrival Rate Estimation

Only a small part of EVs in the TF will go to the FCS on their travel path to receive the charging service, and the number of EVs arrive to the FCS in one time interval is denoted as the average EV arrival rate, λ. As an important input for the charging load forecasting, the FCS charging load cannot be precisely determined if the EV arrival rate is not properly estimated [6][14]. Generally, EV drivers' willingness to go to an FCS on their travel path determines the arrival rate and is highly related to an EV's state-of-charge (SOC) condition, *SOC*, while the charging duration time, μ, is determined by the SOC and the charging power. It is straightforward to understand that the travel distance, D_{ev}, is the key data source used to estimate EV SOCs, which has been applied in [1], [7], [28] and [29]. The relationship between them is briefly represented as

$$SOC=1-p_{ev}D_{ev}/E_{ev} \tag{23}$$

where E_{ev} is an EV's battery capacity; and p_{ev} is an EV's driving consumption power. Due to the variations in EV types, EV parameters and TF data are first normalized to the standard *passenger car unit* (pcu) in this work. The biggest barrier in the current research is that FCSes are still in the primary stage of development, thus real data of travel distance before FCS D_{ev} is difficult to achieve directly. Therefore, the data of *daily travel distance*, D_{ev}^{T}, and *driving in progress distribution*, $g(t)$, which are much more easily obtained from historical data [30][41], are used as the indirect data source for estimating the travel distance of EVs when arriving at the FCS, D_{ev}. From D_{ev}^{T} and $g(t)$, the SOC and arrival rate can be determined.

(1) *Daily travel distance distribution estimation based on the mixture model*: Most of the previously mentioned studies assumed that D_{ev}^{T} follows a simple probability density function (PDF), such as a normal distribution, which does not reflect the actual and cannot reflect the complexity of EV charging behaviors. As shown in Fig. 6, daily travel distance distribution curve is more complicated and several peaks can be seen. To get a more general results, the mixture model based approach is used to analyze the complicated D_{ev}^{T} distribution $p(\mathbf{X}|\mathbf{\Theta})$. The mixture model is able to model random variables using a combination of selected PDFs:

$$p(\mathbf{X}|\mathbf{\Theta})=\sum_{m=1}^{M}\pi_m p_m(x_i\,|\,\theta_m)\qquad 0\leq\pi_m\leq 1,\quad \sum_{m=1}^{M}\pi_m=1 \tag{24}$$

where $\mathbf{X}$ is the random variable set; $\mathbf{\Theta}$ is the PDF parameter; π_m is the weight coefficient of the mth PDF p_m; and M is the PDF number.

The expectation-maximization (EM) algorithm [31] is adopted by using a 2-step iteration to obtain the parameters of the mixture model, which maximizes the expectation of the likelihood:

$$L(\mathbf{\Theta}\,|\,\mathbf{X},\mathbf{Y})=\prod_{i=1}^{N}\sum_{m=1}^{M}y_{m,i}\pi_m p_m(x_i\,|\,\theta_m) \tag{25}$$

where N is the number of data samples; $\mathbf{Y}$ is an auxiliary variable set marking where the data sample belongs to which PDF in the mixture model.

Before applying the EM algorithm, the parameters $\mathbf{\Theta}$ and the auxiliary variables, $\mathbf{Y}$, are initialized using the K-means clustering approach [32]. The PDF number, M, is another critical index that can be obtained using the Akaike information criterion (AIC) and the Bayes information criterion (BIC). These are expressed, respectively, as follows:

$$AIC=2M-2\ln\arg\max_{\mathbf{\Theta}}L(\mathbf{\Theta}|\,\mathbf{X}) \tag{26}$$

$$BIC=M\ln(N)-2\ln\arg\max_{\mathbf{\Theta}}L(\mathbf{\Theta}|\,\mathbf{X}) \tag{27}$$

The optimal M is selected by increasing M one by one until $BIC(M)$-$BIC(M+1)$ and $AIC(M)$-$AIC(M+1)$ are smaller than a pre-set terminating criterion.

The EM algorithm is carried out iteratively, and in the qth iteration, the calculation of the expectation of the likelihood function (25) under the current estimation of $\mathbf{\Theta}^{(q)}$ is the first step:

$$Q(\mathbf{\Theta}\,|\,\mathbf{\Theta}^{(q)})=E[\ln L(\mathbf{\Theta}\,|\,\mathbf{X},\mathbf{Y})\,|\,(\mathbf{Y},\mathbf{\Theta}^{(q)}] \tag{28}$$

Determining the parameters to maximize the expectation in this iteration is the second step:

$$\mathbf{\Theta}^{(q+1)}=\arg\max_{\mathbf{\Theta}}Q(\mathbf{\Theta}|\,\mathbf{\Theta}^{(q)}) \tag{29}$$

The 2-step algorithm is executed iteratively until the pre-set stopping criterion is reached, and the distributions, $p(\mathbf{X})$, are obtained with the parameters optimally estimated using the proposed mixture model. It should be noted the proposed approach is universally applicable and not limited to specific distributions.

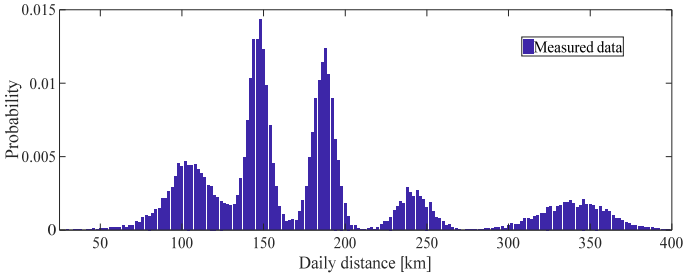

Fig. 6. The distribution of EV daily travel distances.

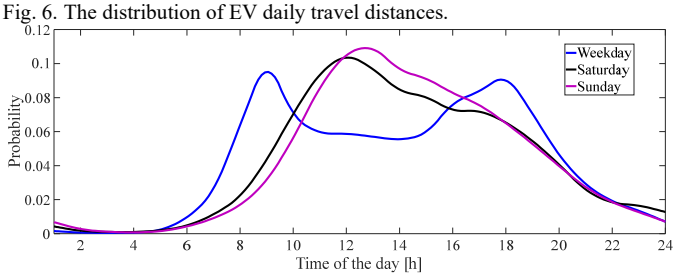

Fig. 7. Trips in progress by time of day in the UK [33].

(2) *Driving in Progress Distribution*: Travel patterns vary with time over the course of a day, and can be represented as the *driving in progress distribution*, $g(t)$. This is also a critical distribution for obtaining the arrival rate at FCSes. As shown in Fig. 7 [33], $g(t)$ is formulated to demonstrate the probability that the EV is currently driving. For instance, $g(5)$ = 8% indicates that the EV has a probability of 8% of being on the road. Based on the definition of $g(t)$, D_{ev}^{T} is estimated using the mean travel distance of EVs at time t, $D_{ev,t}$, where

$$D_{ev}^{T} = D_{ev,t} \Big/ \int_{t_0}^{t} g(h)dh \tag{30}$$

It is assumed that EV drivers prefer to charge when the SOC is below the criterion, $SOC_{\min}$, and their mean SOC is SOC_m ($0 \leqslant SOC_m \leqslant SOC_{\min}$). According to (23) and (30), $D_{ev,t}^{T}$, the daily travel distance of EVs arriving to the FCS D_{ev} at time t is estimated as

$$D_{ev,t}^{T} = (E_{ev}(1-SOC_m)) \Big/ (p_{ev}\int_{t_0}^{t} g(h)dh) \tag{31}$$

The probability that an EV passing by an FCS at time t will choose to charge, P_t, can be estimated using the daily travel distance distribution, $p(\mathbf{X})$:

$$P_t = \int_{D_{ev,t}^{T}}^{D_{ev,t-1}^{T}} p(x)dx \tag{32}$$

Therefore, the arrival rate at time t, λ_t, is:

$$\lambda_t = \varsigma P_t f_t \tag{33}$$

where ς is the penetration rate of EVs during in the total TF; and f_t is the forecasted TF at time t.

B. EV Charging Queuing Model Considering Service Limitations and Driver Behaviors

The capacity of an FCS K is the maximum number of EVs the FCS can accommodate. Currently, most countries are still in the early stages of EV development, and the service ability of FCSes is facing a serious shortage. Meanwhile, many FCS are built in downtown areas or near the expressways. Therefore, the capacity of an FCS is very limited. Thus, the limitations of FCS capacity deserve full consideration. As shown in Fig. 8, an EV charging process is set up as a Markov MMCK model. MMCK is the Kendall notation [8]: It assumes the FCS has C servers and can accommodate a maximum of K EVs; The arrival of EVs is governed by a Markovian-Poisson process [8], in which the arrival EV number, n, in a given time interval follows a Poisson distribution, $P(n)$, with an average EV arrival rate, λ, in each time interval.

$$P(n) = e^{-\lambda}(\lambda^{n}/(n!)) \qquad n \in \Omega^{EV} \tag{34}$$

Also, the charging duration time, t_c, of EVs follows an exponential distribution:

$$f(t_c) = \mu e^{-\mu t_c} \tag{35}$$

where μ is the average EV leave rate for each time interval, which is the number of EVs that finish charging and leave the FCS. According to the nature of exponential distribution, the mean value of charging duration time t_c equals to $1/\mu$. The mean charging duration time t_c depend on the charging levels of the chargers in the FCS as well the *SOC* of EVs et al. Here the subscript t is omitted in order to simplify the expression.

As shown in Fig. 8, sometimes EVs leave the FCS without getting charged. Such EV driver behavior may result from the insufficiency of total number of available chargers, the limitation of FCS capacity, et al., and we categorize the underlying causes of EV leaving behaviors into 3 types: *forced leave*, *refuse to join*, and *impatient leave*. Category 1: *Forced leave*. If the FCS has accommodated K EVs already, other EVs are forced to leave the FCS because there is no space in the queue. Category 2: *Refuse to join*. When arriving to the FCS, the EV finds the waiting line is too long and thus refuses to join the queue, even if the total EV number in the FCS is less than K. Based on categories 1 and 2, the probability that an EV will choose to queue is relevant to the EV number in the FCS, w, and is assumed to be

$$\alpha_w = \begin{cases} 1 & w < C \\ e^{-(w-C)\sigma} & \sigma \geq 0, C \leq w < K \\ 0 & w = K \end{cases} \tag{36}$$

where the exponential function indicates that the probability decreases faster for EVs choosing to queue when w increases, and σ is the parameter that defines the probability decrease rate. Category 3: *Impatient leave*. Even if the waiting line is not too long, the queuing EVs may still choose to leave the FCS when they get too impatient. The number of EVs left in the queue during each time interval is relevant to w and assumed to be

$$b_w = \begin{cases} 0 & w \leq C \\ \delta \ln(w - C + 1) & \delta \geq 0, C < w \leq K \end{cases} \tag{37}$$

where the logarithm function suggests the number of EVs that leave grows as w increases, and δ is the parameter that defines the probability of leaving the FCS. According to (36) and (37), the average arrival rate and the average leaving rate, considering categories 1–3, can be stated as $\lambda_w = \alpha_w\lambda$ and $\mu_w = \mu - b_w$. In practice, the parameters such as impatient leave and refuse to join are usually obtained according to the statistical data, the historical data, surveys or experiments.

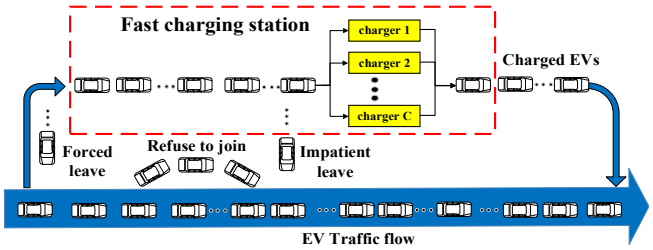

Fig. 8. Capacity constrained queuing model that considers driver behaviors.

C. Stochastic Process Analysis for EV Load Determination

Based on the described queuing model, FCS EV charging loads can be obtained using the stochastic Markov chain analysis. According to stochastic process theory, P_w, the steady-state probability of the Markov chain being in at state w should satisfy the balance equation [34]:

$$\sum_{w\in S} P_w = 1, \quad P_w \sum_{v\in S} q_{wv} = \sum_{v\in S} P_v q_{vw} \qquad \forall w \in S, w \neq v \tag{38}$$

Equation (38) means that the sum of all state probabilities should be 1. Also, the probability in and out of each state should be equal, and q_{wv} is the transition rate from state w to v. S is the state space.

The Markov chain of FCS is shown in Fig. 9. The state transitions in the FCS only belong to the following 2 types: *arrives* and *leaves*. Therefore, based on (38), the FCS balance equation can be expressed as:

$$(\lambda_w + \mu_w) P_w = \lambda_{w-1} P_{w-1} + \mu_{w+1} P_{w+1} \tag{39}$$

By solving (36)–(39), the probability that the FCS has w EVs is:

$$P_w = \begin{cases} (\rho^w / w!)\, P_0 & 0 < w \le C \\ \dfrac{e^{\frac{-(w-C)(w-C-1)}{2}\sigma} \rho^w}{C! \prod_{i=1}^{w-C} [C + \eta \ln(i+1)]} P_0 & C < w \le K \end{cases} \tag{40}$$

The P_0 in (40) can also be obtained by solving (36)–(39):

$$P_0 = \{ \sum_{w=0}^{C} \frac{\rho^w}{w!} + \sum_{w=C+1}^{K} \frac{e^{\frac{-(w-C)(w-C-1)}{2}\sigma} \rho^w}{C! \prod_{i=1}^{w-C} [C + \eta \ln(i+1)]} \}^{-1} \tag{41}$$

where $\rho=\lambda/\mu$ and $\eta=\delta/\mu$ are defined for conciseness. Hence, the charging EV number in an FCS can be expressed as

$$\begin{aligned} N_{ch} &= \sum_{w=0}^{C} w P_w + C \sum_{w=C+1}^{K} P_w \\ &= \sum_{w=0}^{C} w \frac{\rho^w}{w!} P_0 + C \sum_{w=C+1}^{K} \frac{e^{\frac{-(w-C)(w-C-1)}{2}\sigma} \rho^w}{C! \prod_{i=1}^{w-C} [C + \eta \ln(i+1)]} P_0 \end{aligned} \tag{42}$$

and the charging load of the FCS, P^{FCS}, can be determined using N_{ch} and the power of the charging post, p^{EV}:

$$P^{FCS} = N^{ch} p^{EV} \tag{43}$$

And the values of p^{EV} are different according to different charging levels of the chargers.

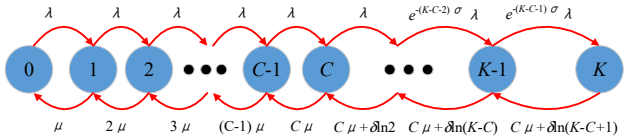

Fig. 9. State transition diagram of the Markov chain considering EV behaviors.

IV. Case Study

A. Simulation Setting Description

The proposed deep learning approach is tested using TF data from the M42 motorway between sites J5 and J6 (GPS Ref: 417049:278576-419805:283048) in England, UK [35]. The TF data with 15 minute temporal resolutions from Jan to Dec 2014 are analyzed. Factors such as weather conditions and travel purposes vary significantly during different days or seasons and lead to prominent uncertainties in TF. By considering the seasonal and weekly differences and diversity, the proposed approach is tested using TF data from weekdays and weekends during the 4 different seasons. The data for the first 2 months of each season are chosen as the training dataset, and the remaining 1 month of data are chosen as the testing dataset. The input data are decomposed using the WT approach into 1 approximation frequency and 3 detail frequencies. The structure of CNN consisted of 4 convolution layers, 2 pooling layers, 1 fully connected layer and 1 output layer for each frequency. The utilized hourly TF data is sampled in 15-min intervals. Four 15-min intervals over 1 hour are averaged to get the hourly TF data used in this work, and 36 hourly TF data is the input of CNN for prediction. The obtained results are compared using the back-propagation neural network (BPNN), support vector machine (SVM), SAE, time-delayed neural network (TDNN), growing deep belief network (DBN) and recurrent neural network (RNN) approaches. All the methods are trained and tested after conducting the wavelets.

Parameters and hyper parameters of different approaches used in comparisons should be properly designed to obtain good forecasting results for comparison. These parameters are chosen by the validation test in a trail-and-error manner. For example, as shown in Fig. 10 and Fig. 11, the performance of BPNN and SAE is assessed by the MAPE. 41 hidden neurons are chosen by the BPNN to reach the smallest MAPE, and the SAE with 5 hidden layers can ensure the optimal MAPE. Kernel function type is critical for the performance of SVM. As shown in TABLE I, comparing with other typical kernel functions such as linear, polynomial and sigmoid kernel functions, the SVM with radial basis function (RBF) has the best MAE, RMSE and MAPE, and RBF is chosen as the kernel function.

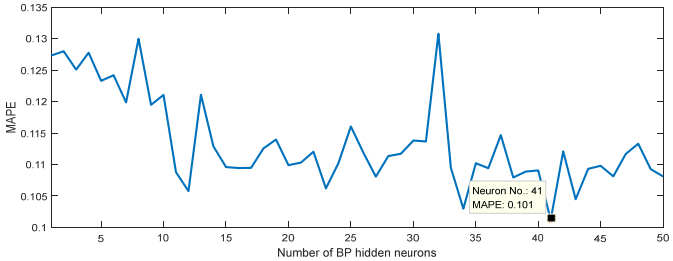

Fig. 10. Validation test for BPNN with different number of hidden neurons.

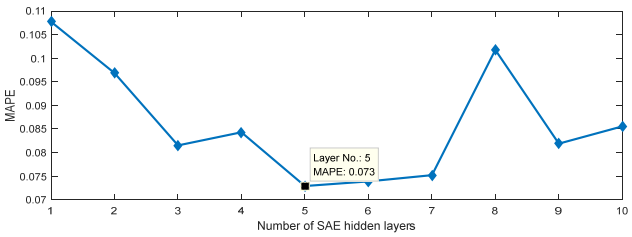

Fig. 11. Validation test for SAE with different number of hidden layers.

TABLE I
VALIDATION TEST FOR SVM WITH DIFFERENT KERNEL FUNCTION TYPES

Kernel function type	MAE	RMSE	MAPE
Linear	363.7714	506.7461	13.29%
Polynomial	534.6109	739.4766	19.21%
RBF	**228.1530**	**348.2298**	**8.20%**
Sigmoid	272.3463	360.3255	9.45%

B. TF Forecasting Results Analysis

The seasonal deterministic 1-hour ahead forecast absolute residuals of the different approaches are shown in Fig. 12 and Fig. 13. *Residual* is the difference between the actual value and the forecasted value. It can be observed from the two figures that the forecasted results of the proposed approach have the smallest residuals, indicating that, among the 6 approaches, the forecasted TF curve is closest to the actual TF curve. Therefore, the comparative results visually demonstrate the effective forecast capability of the proposed approach. The results can be explained by the greater ability of the CNN approach to extract the complexity and non-smoothness of the TF series. Forecast performance evaluation indices are illustrated in TABLE II. The indices of MAE, RMSE and MAPE during the different seasons in TABLE II also show that the errors of the proposed approach are approximately half those obtained using other approaches, indicating that the CNN provided the best point forecast performance over the benchmarks.

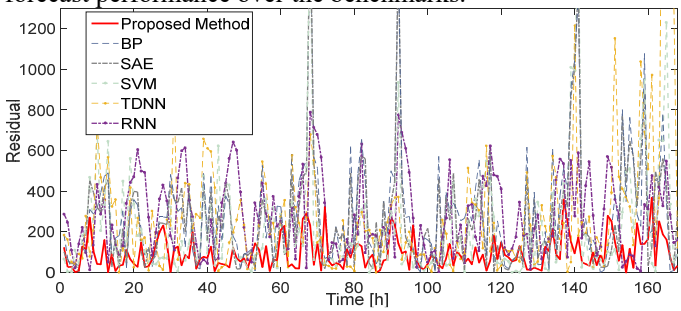

Fig. 12. Deterministic TF forecast absolute residual in summer at J5-J6 M42.

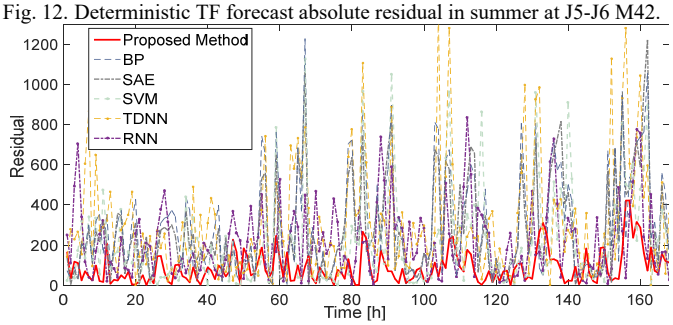

Fig. 13. Deterministic TF forecast absolute residual in winter at J5-J6 M42.

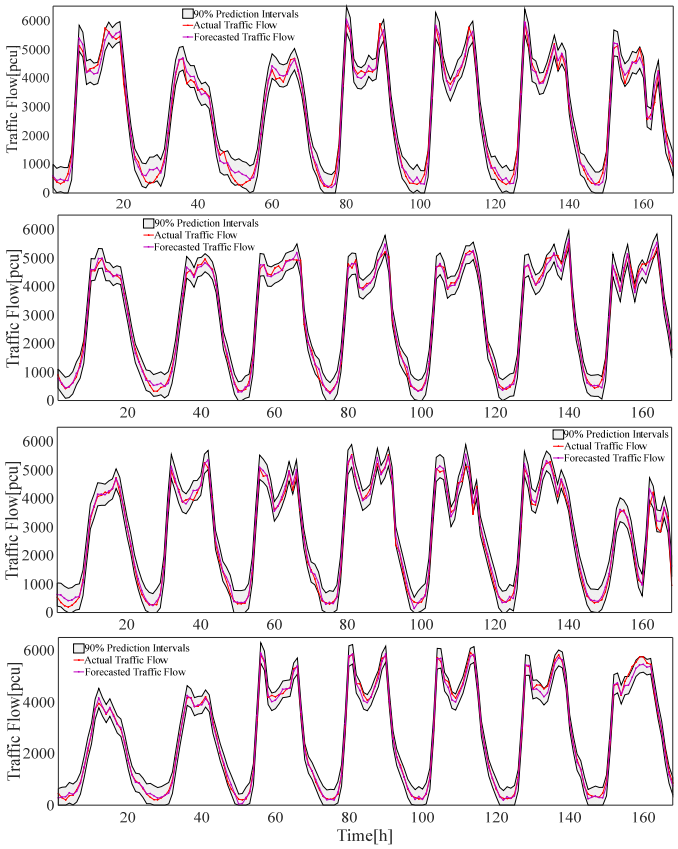

Fig. 14. Probabilistic one-hour ahead TF forecast results during the spring, summer, fall, and winter of 2014.

Compared with the deterministic forecast, probabilistic forecast evaluates the impacts of data uncertainty and model uncertainty in the forecast procedures. PICP, ACE, PL, and IS, as shown in TABLE III, are used as the indices to evaluate the probabilistic performance. PI nominal confidence (PINC) levels of 90%, 95% and 99% are used since high reliability is desired for FCS operation and control. The PICP is supposed to be the closest to PINC for reliable PIs. Meanwhile, the difference between the PICP and PINC, which is defined as ACE, is supposed to be as close to zero as possible. It can be seen from the PICP index that the proposed approach has the most reliable PIs of the measured TF. For example, at the confidence level with PINC = 95%, the CNN approach generates PICPs of 95.49%, 95.24%, 95.47% and 95.24% for 4 seasons, and outperform all other approaches. Although the PICP is only a bit better than that of other approaches, the indices of ACE and IS are only half those of the other approaches, and the improvements are significant. In all 4 seasons, the ACEs of the proposed approach are smaller than 1%, showing the lowest deviations from the nominal confidence levels, and prove the least reliability errors and highest forecast reliability among the benchmarks. A small absolute IS means higher interval sharpness, and the IS index in TABLE III also indicates that the proposed approach outperformed the other 5 approaches from the perspective of interval sharpness. E.g., the CNN approach generates PIs with absolute IS index of 28.64 and 23.91 in spring and summer at the confidence level 99%, which are the smallest compared to all other approaches. The lower PL indicates the more accurate probabilistic forecast results. It can be observed in TABLE III

that the CNN has the lowest scoring metric of PL, which suggest the best accuracy performance among all the methods.

Fig. 14 shows the actual TF, the forecasted TF, as well as the established PIs with PINC 90% for the 4 seasons. It is obvious that the curves in the 4 graphs fluctuate widely, indicating the nonlinear and non-stationary features of TF data during the 4 seasons. It can be seen that in each graph, the shapes of the lower and upper bounds, as well as the actual and forecasted TF curves, are very similar to each other. This shows that the actual and forecasted TF are perfectly enclosed by the constructed upper and lower bound, indicating that the probabilistic performance of the proposed approach is satisfactory for the construction of high-performance PIs. Furthermore, the proposed probabilistic approach is compared with the well-known quantile regression (QR) method to verify its performance. In Fig. 15, The ACE deviations and PL values of the proposed approach are small than QR benchmarks in all 4 seasons with the PINC 90%, 95% and 99%, respectively, which indicates the proposed approach performs better than the QR benchmarks with respect to 4 seasons and 3PINCs.

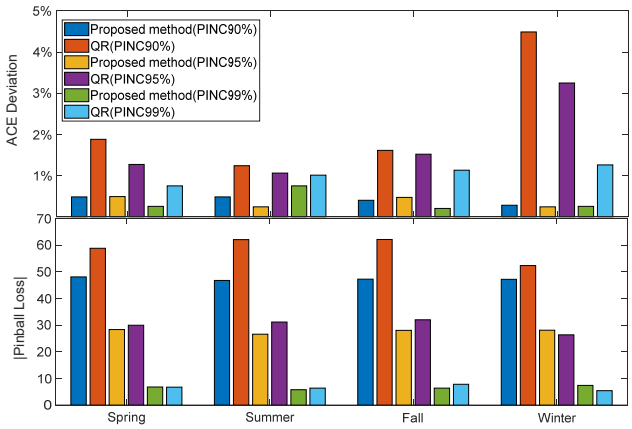

Fig. 15. ACE and PL of QR and the proposed approach in different seasons.

The proposed prediction method is capable of different time scales. For example, when day-ahead decisions are involved, 24 hour-ahead prediction should be conducted. To effectively handle this requirement, the utilized CNN is adjusted from single-output to multiple-output (24-output for day-ahead prediction), and the results of 24 hour-ahead prediction are obtained. Fig. 16 gives the result of 24-hour ahead TF forecasting, which could be used in day-ahead decisions. Meanwhile, Table IV gives the probabilistic 6-hour, 12-hour, 18-hour and 24-hour ahead TF forecasting errors. Probabilistic TF forecast errors with different time horizon are shown in Fig. 17. The results demonstrate the proposed approach could reliably handle the time inconsistency.

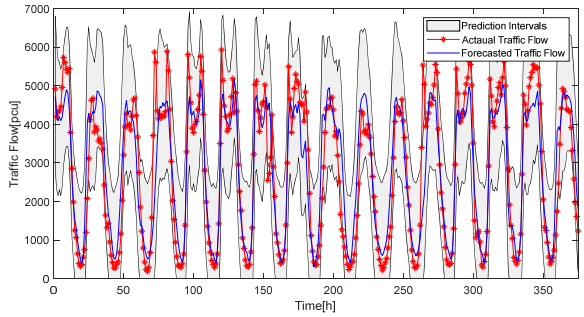

Fig. 16. Probabilistic 24-hour ahead TF forecast results during the spring.

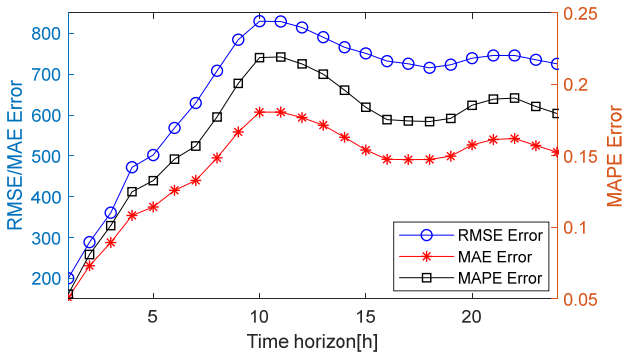

Fig. 17. Deterministic TF forecast errors with different time horizon.

TABLE IV
6-hour, 12-hour and 18-hour 24-hour Ahead TF Forecasting Error

	Index	Spring		
6h Ahead	PINC	0.9	0.95	0.99
	PICP	96.00%	97.60%	99.20%
	ACE	6.00%	2.60%	0.20%
	IS	-610.711	-358.459	-88.7261
	PL	143.026	84.41546	21.5182
12h Ahead	Index	Spring		
	PINC	0.9	0.95	0.99
	PICP	95.73%	97.60%	99.20%
	ACE	5.73%	2.60%	0.20%
	IS	-772.385	-452.996	-114.866
	PL	181.8217	107.2718	27.69189
18h Ahead	Index	Spring		
	PINC	0.9	0.95	0.99
	PICP	96.00%	98.13%	98.93%
	ACE	6.00%	3.13%	-0.07%
	IS	-814.115	-492.194	-146.422
	PL	188.9601	113.3132	31.93704
24h Ahead	Index	Spring		
	PINC	0.9	0.95	0.99
	PICP	96.00%	97.87%	99.47%
	ACE	6.00%	2.87%	0.47%
	IS	-769.264	-452.774	-122.986
	PL	180.9664	106.9679	28.63418

C. Charging Load Forecasting Results Analysis

In the simulation, the FCS has the rated power of $P_r^{FCS} = 0.88$ MW combined with $C = 22$ chargers of $p^{EV} = 40$ KW and could accommodate a maximum of $K = 30$ EVs. The average charging duration time is $t_c = 20$ minutes. The default parameters of EVs are set to be $\varsigma = 20\%$, $\sigma = 1$ and $\delta = 1$.

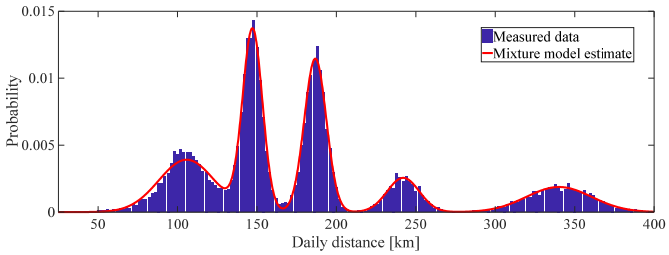

Fig. 18. The PDF of EV daily travel distances estimated by the mixture model.

The distribution of EV daily travel distances are estimated using the mixture model, in which the normal distributions are chosen as the selected PDFs; and the results are shown in Fig. 18. It is clear that the curve estimated using the mixture model closely follows the trend of the actual data, which graphically confirms the effectiveness of the proposed model.

Based on the TF forecast results and the established model in Section. III, the forecasted charging load and the 90% confidence PIs during the different seasons are visually presented in Fig. 19, which demonstrates that the trend of the charging load curve is not the same as the TF and fluctuates more significantly. This is due to the significant stochasticity of EV charging behavior, as well as the nonlinearity and complexity of the charging load dataset. It is also found that the PI of charging load is much narrower than that of the TF. This is due to the FCS parameters, as well as to the nonlinear relationship of charging load and the influence of TF on the construction of charging load PIs. However, the forecasted curve is still within the constructed PIs by a large percentage. Therefore, it can be concluded the probabilistic performance of the proposed approach is satisfactory for the operation and control of FCS. TABLE V and TABLE VI give the deterministic and probabilistic one-hour ahead charging load forecasting errors respectively. Results in TABLE V indicate the forecasting performance are acceptable in all 4 seasons. Meanwhile, the results in TABLE VI verify that the narrower PI of charging load leads to good probabilistic forecasting errors with the 90% PINC.

TABLE V
DETERMINISTIC ONE-HOUR AHEAD CHARGING LOAD FORECASTING ERROR

Season	Error Index	Error Value
Spring	MAE	11.29683
	RMSE	19.30667
	MAPE	2.62%
Summer	MAE	12.94862
	RMSE	20.37905
	MAPE	2.90%
Fall	MAE	14.65326
	RMSE	24.07438
	MAPE	3.52%
Winter	MAE	16.23156
	RMSE	26.24738
	MAPE	3.78%

TABLE VI
PROBABILISTIC ONE-HOUR AHEAD CHARGING LOAD FORECASTING ERROR

	Spring	Summer	Fall	Winter
PINC	90%	90%	90%	90%
PICP	99.40%	100%	98.81%	96.43%
ACE	9.40%	10%	8.81%	6.43%
PL	7.510393	7.318551	7.508211	7.629631
IS	-30.5677	-29.2742	-30.9993	-31.2726

To further explore the trend differences between TF curve and charging load curve as shown in Fig. 14 and Fig. 19, their correlations are studied. TABLE VII gives the Pearson (linear), Kendall, and Spearman correlation coefficients [36] of TF and charging load during the different seasons. All the coefficients range between 0.8 to 1 and show that, although there is a strong relationship between the 2 time series, the trend differences should not be neglected. Fig. 20 and Fig. 21 demonstrate the Pearson (linear) correlation and polynomial fitting between TF and charging load, which further clarifies the increasing monotonic trend between TF and charging load. However, this relationship is neither strictly linear nor strictly polynomial fitting due to the complicated nonlinear transformation associated with the proposed queuing model.

TABLE VII
CORRELATION COEFFICIENTS OF TF AND CHARGING LOAD

Season	Pearson	Kendall	Spearman
Spring	0.9414	0.8041	0.9490
Summer	0.9321	0.8298	0.9373
Fall	0.9274	0.8303	0.9504
Winter	0.9533	0.8455	0.9584

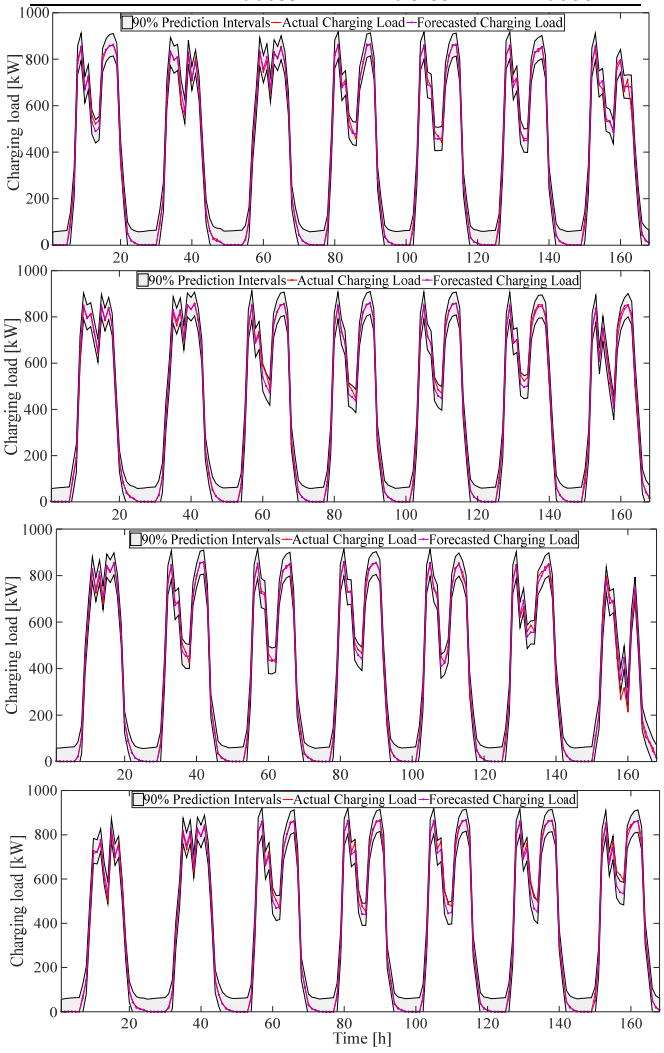

Fig. 19. Probabilistic one-hour ahead charging load forecast results during the spring, summer, fall, and winter of 2014.

It is assumed in this study that driver behaviors have a significant influence on the charging load forecast results, and are classified roughly as forced leave, refuse to join, and impatient leave. Fig. 22 shows that the charging load becomes larger when the charging post number, C, or capacity, K, of the FCS increases, since fewer EVs will be forced to leave due to limitations of FCS service capability. Clearly, the shortages of

charging posts and charging capacity impose restrictions on the wide use of EVs, and the installation of enough charging facilities, which can effectively enhance EV driver charging convenience, will make EVs more popular. Driver sensitivity to charging wait time is another factor that can influence the charging load. Fig. 23 demonstrates that if drivers are more sensitive to waiting time, which translates to a larger "refuse to join" probability parameter, σ, or a larger "impatient leave" probability parameter, δ, the charging load decreases accordingly. From the curves of Fig. 23, it can also be seen that the charging load reduction becomes particularly evident during rush hour, when there is higher charging demand. This indicates that flexible strategies should be taken in the FCS operation process to promote service capabilities, such as the establishment of time-of-use prices for EV charging load, et al.

V. Conclusions

The established prediction approach for EV charging demand that has better accuracy and reliability is essential for the operation and control of FCSes, the power system, and the traffic system. Such an approach will help FCS operators determine the number of dispatchable EVs in advance to avoid higher costs or excessive risks. This will also assist the traffic operator in the alleviation of traffic congestion and will aid decision-making for electricity market trading using optimized bidding when EVs participate in the ancillary service market.

TABLE II
Deterministic One-hour Ahead TF Forecasting Error

Season	Error	BPNN	SVM	SAE	TDNN	RNN	DBN	**CNN**
Spring	MAE	296.5926	229.8162	265.6717	359.9786	260.8728	278.5046	93.4223
	RMSE	416.6938	341.5507	382.4075	535.3427	331.3600	396.0828	118.3225
	MAPE	10.56%	8.18%	9.46%	12.87%	9.23%	9.92%	3.29%
Summer	MAE	291.2690	238.1329	271.2713	291.2655	267.3694	264.135	94.2701
	RMSE	398.1834	369.2117	382.9356	433.1820	324.6583	378.6133	120.6187
	MAPE	10.08%	8.24%	9.39%	10.14%	9.20%	9.14%	3.26%
Fall	MAE	337.7184	263.1371	312.5201	336.8291	307.9589	322.0832	107.1443
	RMSE	482.9903	412.7680	464.0448	498.2443	371.3890	474.9173	140.9987
	MAPE	12.61%	9.83%	11.67%	12.63%	11.46%	12.03%	3.97%
Winter	MAE	273.3713	181.5259	258.1541	291.2980	232.2863	255.0828	83.3157
	RMSE	372.9150	269.3890	351.7644	392.0209	297.3276	343.3366	110.2241
	MAPE	9.88%	6.56%	9.33%	10.61%	8.33%	9.22%	3.05%

TABLE III
Probabilistic One-hour Ahead TF Forecasting Error

Season		Spring			Summer			Fall			Winter		
	PINC	90%	95%	99%	90%	95%	99%	90%	95%	99%	90%	95%	99%
BPNN	PICP	87.50%	91.11%	94.44%	87.50%	92.22%	96.11%	88.61%	92.22%	96.11%	88.06%	92.50%	95.83%
	ACE	-2.50%	-3.89%	-4.56%	-2.50%	-2.78%	-2.89%	-1.39%	-2.78%	-2.89%	-1.94%	-2.50%	-3.17%
	PL	76.50	51.59	16.19	73.61	49.71	16.12	85.52	55.70	17.81	75.44	46.95	22.51
	IS	-392.53	-254.72	-94.43	-353.59	-233.02	-96.97	-412.37	-266.90	-104.85	-557.25	-340.52	-109.12
SVM	PICP	97.50%	98.61%	99.72%	96.11%	97.78%	99.44%	90.56%	94.44%	97.22%	99.44%	100%	100%
	ACE	7.50%	3.61%	0.72%	6.11%	2.78%	0.44%	0.56%	0.56%	1.78%	9.44%	5.00%	1.00%
	PL	102.34	60.50	15.70	102.37	60.46	15.77	89.20	53.58	15.23	100.85	59.78	15.74
	IS	-391.40	-242.33	-63.37	-429.39	-252.26	-65.27	-398.32	-241.41	-72.52	-404.19	-239.13	-62.95
SAE	PICP	88.89%	93.06%	96.67%	90.83%	94.17%	96.94%	89.17%	92.50%	96.11%	85.00%	90.56%	96.94%
	ACE	-1.11%	-1.94%	-2.33%	0.83%	-0.83%	-2.06%	-0.83%	-2.50%	-2.89%	-5.00%	-4.44%	-2.06%
	PL	78.18	48.29	15.90	76.80	53.00	13.42	91.26	55.90	19.04	70.75	42.62	11.65
	IS	-383.54	-240.65	-87.73	-345.91	-215.25	-70.13	-449.23	-286.78	-108.37	-341.63	-199.79	-58.14
TDNN	PICP	92.72%	94.40%	97.48%	91.04%	94.40%	96.64%	91.88%	93.00%	96.36%	91.32%	94.68%	96.36%
	ACE	2.72%	-0.60%	-1.52%	1.04%	-0.60%	-2.36	1.88%	-2.00%	-2.64%	1.32%	-0.32%	-2.64%
	PL	110.87	67.26	19.76	73.80	45.37	13.66	86.74	53.51	17.92	73.38	44.91	13.70
	IS	-526.28	-322.06	-89.36	-353.28	-212.99	-63.12	-421.49	-264.43	-89.70	-360.11	-232.58	-91.14
RNN	PICP	93.06%	96.39%	97.78%	86.11%	91.67%	95.83%	86.67%	90.83%	96.67%	88.61%	93.89%	98.61%
	ACE	3.06%	1.39%	-1.22%	-3.89%	-3.33%	-3.17%	-3.33%	-4.17%	-2.33%	-1.39%	-1.11%	-0.39%
	PL	52.56	32.33	9.63	50.64	31.70	11.22	53.51	33.12	9.68	46.94	28.13	7.78
	IS	-226.49	-140.14	-44.03	-232.61	-148.19	-60.35	-248.07	-154.94	-46.74	-209.69	-123.70	-33.97
DBN	PICP	91.39%	95.56%	98.33%	91.11%	95.83%	98.33%	90.56%	93.61%	97.78%	90.83%	93.61%	98.89%
	ACE	1.39%	0.56%	-0.67%	1.11%	0.83%	-0.67%	0.56%	-1.39%	-1.22%	0.83%	-1.39%	-0.11%
	PL	98.60	57.99	15.76	84.24	50.82	13.93	87.04	53.12	17.08	80.39	50.25	12.59
	IS	-444.08	-259.24	-72.19	-386.86	-236.04	-82.564	-455.51	-269.22	-84.64	-389.71	-223.64	-53.79
CNN	PICP	90.48%	95.49%	99.25%	90.48%	95.24%	98.25%	90.40%	95.47%	99.20%	89.72%	95.24%	98.75%
	ACE	0.48%	0.49%	0.25%	0.48%	0.24%	-0.75%	0.40%	0.47%	0.20%	-0.28%	0.24%	-0.25%
	PL	48.10	28.32	6.76	46.78	26.60	5.78	47.24	28.03	6.34	47.20	28.06	7.36
	IS	-176.06	-102.68	-28.64	-191.72	-108.51	-23.91	-170.34	-99.94	-27.05	-201.04	-121.04	-30.64

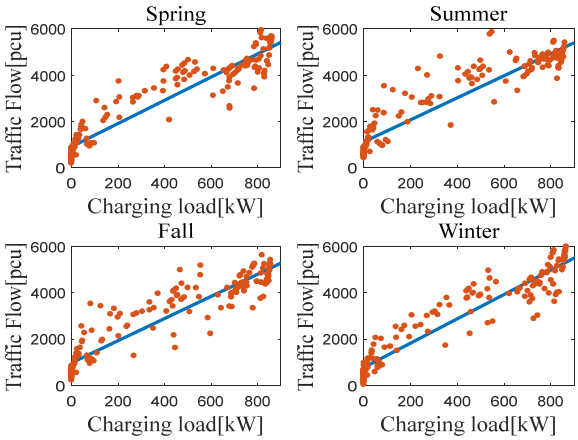

Fig. 20. The Pearson (linear) correlation between TF and charging load during different seasons.

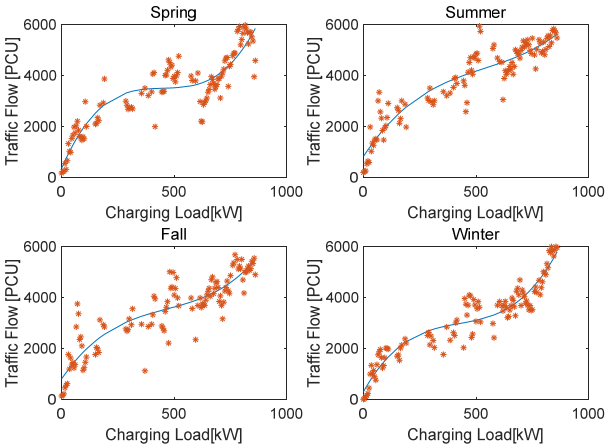

Fig. 21. Polynomial fitting between the TF and the charging load.

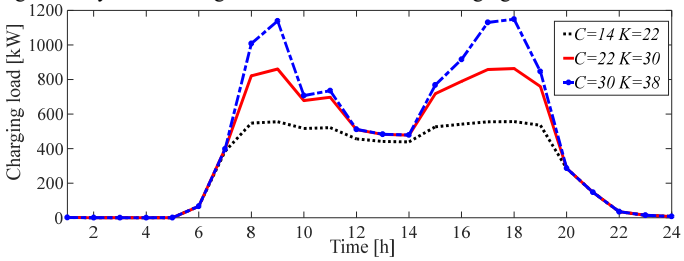

Fig. 22. Forecasted charging load for different values of C and K.

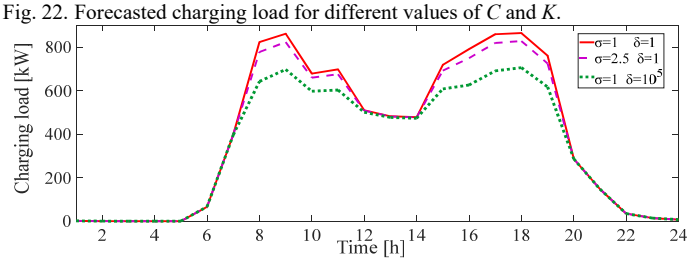

Fig. 23. Forecasted charging load for different values of σ and δ.

VI. References

[1] G. Wang *et al*., "Robust Planning of Electric Vehicle Charging Facilities With an Advanced Evaluation Method," *IEEE Trans. Ind. Informat.*, vol. 14, no. 3, pp. 866-876, Mar. 2018.

[2] G. Wang, J. Zhao, F. Wen, Y. Xue and G. Ledwich, "Dispatch Strategy of PHEVs to Mitigate Selected Patterns of Seasonally Varying Outputs from Renewable Generation," *IEEE Trans. Smart Grid*, vol. 6, no. 2, pp. 627-639, Mar. 2015.

[3] Z. Duan, B. Gutierrez and L. Wang, "Forecasting Plug-In Electric Vehicle Sales and the Diurnal Recharging Load Curve," *IEEE Trans. Smart Grid*, vol. 5, no. 1, pp. 527-535, Jan. 2014.

[4] N. Korolko, Z. Sahinoglu and D. Nikovski, "Modeling and Forecasting Self-Similar Power Load Due to EV Fast Chargers," *IEEE Trans. Smart Grid*, vol. 7, no. 3, pp. 1620-1629, May 2016.

[5] Q. Dai, T. Cai, S. Duan and F. Zhao, "Stochastic Modeling and Forecasting of Load Demand for Electric Bus Battery-Swap Station," *IEEE Trans. Power Del.*, vol. 29, no. 4, pp. 1909-1917, Aug. 2014.

[6] M. Alizadeh, A. Scaglione, J. Davies and K. S. Kurani, "A scalable stochastic model for the electricity demand of electric and plug-In hybrid vehicles," *IEEE Trans. Smart Grid*, vol. 5, no. 2, pp. 848-860, Mar. 2014.

[7] D. Wu, D. C. Aliprantis and K. Gkritza, "Electric Energy and Power Consumption by Light-Duty Plug-In Electric Vehicles," *IEEE Trans. Power Syst.*, vol. 26, no. 2, pp. 738-746, May 2011.

[8] A.Y. Khinchine, *Mathematical Methods in the Theory of Queueing*, New York: Dover Publications, 1960.

[9] A. Y. S. Lam, K. C. Leung and V. O. K. Li, "Capacity estimation for vehicle-to-grid frequency regulation services with smart charging mechanism," *IEEE Trans. Smart Grid*, vol. 7, no. 1, pp. 156-166, Jan. 2016.

[10] P. Fan, B. Sainbayar and S. Ren, "Operation Analysis of Fast Charging Stations with Energy Demand Control of Electric Vehicles," *IEEE Trans. Smart Grid*, vol. 6, no. 4, pp. 1819-1826, Jul. 2015.

[11] U. C. Chukwu and S. M. Mahajan, "V2G Parking Lot with PV Rooftop for Capacity Enhancement of a Distribution System," *IEEE Trans. Sustain. Energy.*, vol. 5, no. 1, pp. 119-127, Jan. 2014.

[12] Y. Sheffi, *Urban Transportation Network: Equilibrium Analysis with Mathematical Programming Methods*. Upper Saddle River, NJ, USA: Prentice Hall, 1985.

[13] S. Bae and A. Kwasinski, "Spatial and Temporal Model of Electric Vehicle Charging Demand," *IEEE Trans. Smart Grid*, vol. 3, no. 1, pp. 394-403, Mar. 2012.

[14] X. Dong, Y. Mu, H. Jia, J. Wu and X. Yu, "Planning of Fast EV Charging Stations on a Round Freeway," *IEEE Trans. Sustain. Energy.*, vol. 7, no. 4, pp. 1452-1461, Oct. 2016.

[15] B. Ghosh, B. Basu and M. O'Mahony, "Multivariate Short-Term Traffic Flow Forecasting Using Time-Series Analysis," *IEEE Trans. Intell. Transp. Syst.*, vol. 10, no. 2, pp. 246-254, Jun. 2009.

[16] S. Oh, Y. Byon and H. Yeo, "Improvement of Search Strategy with K-Nearest Neighbors Approach for Traffic State Prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 4, pp. 1146-1156, April 2016.

[17] Y. Hou, P. Edara and C. Sun, "Traffic flow forecasting for urban work zones," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 4, pp. 1761-1770, Aug. 2015.

[18] M. T. Sánchez-Rico, R. García-Ródenas and J. L. Espinosa-Aranda, "A Monte Carlo Approach to Simulate the Stochastic Demand in a Continuous Dynamic Traffic Network Loading Problem," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 3, pp. 1362-1373, Jun. 2014.

[19] H. F. Yang, T. S. Dillon and Y. P. P. Chen, "Optimized structure of the traffic flow forecasting model with a deep learning approach," *IEEE Trans. Neu. Netw. Learn. Syst*, vol. 28, no. 10, pp. 2371-2381, Oct. 2017.

[20] A. Koesdwiady, R. Soua and F. Karray, "Improving Traffic Flow Prediction with Weather Information in Connected Cars: A Deep Learning Approach," *IEEE Trans. Veh. Technol*, vol. 65, no. 12, pp. 9508-9517, Dec. 2016.

[21] W. Huang, G. Song, H. Hong and K. Xie, "Deep Architecture for Traffic Flow Prediction: Deep Belief Networks with Multitask Learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 5, pp. 2191-2201, Oct. 2014.

[22] Y. Lv, Y. Duan, W. Kang, Z. Li and F. Y. Wang, "Traffic Flow Prediction with Big Data: A Deep Learning Approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 2, pp. 865-873, Apr. 2015.

[23] O. Abdel-Hamid, A. r. Mohamed, H. Jiang, L. Deng, G. Penn and D. Yu, "Convolutional Neural Networks for Speech Recognition," *IEEE/ACM Trans. Audio, Speech, and Language Processing*, vol. 22, no. 10, pp. 1533-1545, Oct. 2014.

[24] Y. Liang, J. Wang, S. Zhou, Y. Gong, N. Zheng, "Incorporating image priors with deep convolutional neural networks for image super-resolution," *Neurocomp.*, vol. 194, no. 19, pp: 340-347, Jun. 2016.

[25] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks." *International Conference on Neural Information Processing Systems*, Curran Associates Inc. vol. 25, pp: 1097-1105, 2012.

[26] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674-693, Jul. 1989.

[27] H. Cramér, *Mathematical Methods of Statistics*, Princeton, NJ, USA: Princeton Univ., 1971.

[28] G. Wang, Z. Xu, F. Wen and K. P. Wong, "Traffic-Constrained Multiobjective Planning of Electric-Vehicle Charging Stations," *IEEE Trans. Power Del.*, vol. 28, no. 4, pp. 2363-2372, Oct. 2013.

[29] M. J. Mirzaei, A. Kazemi and O. Homaee, "A Probabilistic Approach to Determine Optimal Capacity and Location of Electric Vehicles Parking Lots in Distribution Networks," *IEEE Trans. Ind. Informat.*, vol. 12, no. 5, pp. 1963-1972, Oct. 2016.

[30] S. Negarestani, M. Fotuhi-Firuzabad, M. Rastegar and A. Rajabi-Ghahnavieh, "Optimal Sizing of Storage System in a Fast Charging Station for Plug-in Hybrid Electric Vehicles," *IEEE Trans. Transport. Electrific.*, vol. 2, no. 4, pp. 443-453, Dec. 2016.

[31] Dempster, A. P., et al. "Maximum Likelihood from Incomplete Data via the EM Algorithm." *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, no. 1, pp. 1–38, 1977.

[32] C. Carson, S. Belongie, H. Greenspan and J. Malik, "Blobworld: image segmentation using expectation-maximization and its application to image querying," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 8, pp. 1026-1038, Aug. 2002.

[33] UK Department for Transport, *National Travel Survey*. [Online]. Available:https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/79153/nts0501.xls

[34] F. Kelly, *Reversibility and stochastic networks*, Cambridge University Press New York, NY, USA: 2011.

[35] Highways England, *Traffic Flow Data*, [Online]. Available: http://tris.highwaysengland.co.uk/download/6e8f2a60-e4bc-4805-a5d2-d7f5c4a992db

[36] M. Hollander, D. Wolfe, and E. Chicken, *Nonparametric Statistical Methods*, Hoboken, NJ, USA: Wiley, 2014.

[37] W. Wei, S. Mei, L. Wu, J. Wang and Y. Fang, "Robust operation of distribution networks coupled with urban transportation infrastructures," *IEEE Trans. Power Syst*., vol. 32, no. 3, pp. 2118-2130, May 2017.

[38] W. Wei, L. Wu, J. Wang and S. Mei, "Network Equilibrium of Coupled Transportation and Power Distribution Systems," *IEEE Trans. Smart Grid*, vol. PP, no. 1, pp. 1-1, Early access 2018.

[39] W. Wei, L. Wu, J. Wang and S. Mei, "Expansion Planning of Urban Electrified Transportation Networks: A Mixed-Integer Convex Programming Approach," *IEEE Trans. Transport. Electrific.*, vol. 3, no. 1, pp. 210-224, March 2017.

[40] C. M. Bishop, *Neural Networks for Pattern Recognition*, London, U. K.: Oxford Univ. Press, 1996.

[41] A. Santos, N. McGuckin, H. Y. Nakamoto, D. Gay, and S. Liss, Summary of Travel Trends: 2009 National Household Travel Survey, U.S. Department of Transportation Federal Highway Administration, Washington, DC, USA, Rep. FHWA-PL-11022, Jun. 2011.

[42] J. Bi, H. Yuan, L. B. Zhang, and J. Zhang, "SGW-SCN: "An Integrated Machine Learning Approach for Workload Forecasting in Geo-Distributed Cloud Data Centers," *Information Sciences*, vol. 481, pp. 57-68, May 2019.

[43] J. Tang, C. Deng, G. B. Huang, "Extreme learning machine for multilayer perceptron," *IEEE Transactions on Neural Networks and Learning Systems*, vol 27, no. 4, pp. 809-821, 2015.

[44] G. M. Wang, J. F. Qiao, J. Bi, W. J. Li, and M. C. Zhou, "TL-GDBN: Growing Deep Belief Network with Transfer Learning," *IEEE Transactions on Automation Science and Engineering*, vol. 16, no. 2, pp. 874-885, 2018.

[45] C. L. P. Chen, Z. Liu, "Broad learning system: An effective and efficient incremental learning system without the need for deep architecture," *IEEE transactions on neural networks and learning systems*, vol. 29, no. 1, pp. 10-24, 2017.

[46] Y. LeCun, "LeNet-5, convolutional neural networks," [Online], Available: http://yann.lecun.com/exdb/lenet/

[47] J. Tanida, et al. "Parallel distributed processing model with local space-invariant interconnections and its optical architecture." *Applied Optics*, vol. 29, no. 32, pp. 4790-4797, Nov. 1990.