Extraction of Spatial-Temporal Features of Bus Loads in Electric Grids Through Clustering in a Dynamic Model Space

WEI ZHANG1,2, GANG MU1,2, CHEN SONG3,4, GANGUI YAN1,2, AND VINCENT HEUVELINE3,4

1School of Electrical Engineering, Northeast Electric Power University, Jilin 132012, China
2Key Laboratory of Modern Power System Simulation and Control & Renewable Energy Technology, Ministry of Education, Northeast Electric Power University, Jilin 132012, China
3Engineering Mathematics and Computing Lab (EMCL), Interdisciplinary Center for Scientific Computing (IWR), Heidelberg University, 69120 Heidelberg, Germany
4Heidelberg Institute for Theoretical Studies, 69118 Heidelberg, Germany

Corresponding author: Gang Mu (mg@neepu.edu.cn)

This work was supported in part by the National Nature Science Foundation of China under Grant NSFC-51437003, and in part by the Heidelberg Institute for Theoretical Studies.

ABSTRACT Bus loads in electric grids have inherently a spatial-temporal behavior and also a certain degree of randomness. The spatial-temporal feature based bus load forecasting, which provides additional information on the spatial distribution and the uncertainty of future electric loads, is of importance to power systems dispatching and planning, in particular, with intermittent renewable power generation. In this paper, a method for extracting spatial-temporal features, including abnormal states of multiple bus loads in electric grids, is proposed. The abnormal spatial load states are firstly identified by using one-class support vector machine. Then, only the load fluctuations of normal states are mapped into a dynamic model space supported by polynomials in order to approximate the time series of bus loads. The parameters of polynomials are clustered by the Dirichlet process mixture model for deriving the patterns of load state evolution. As a result, the extracted spatial-temporal patterns are a set of different distributions of bus loads with static features and dynamic features displayed explicitly. The method is tested against the bus loads of an electric grid in a city in the Northeast China. The proposed methodology is validated with respect to the bus loads in time slots of the future 10 days.

INDEX TERMS Load forecasting, multiple bus loads, one-class support vector machine, spatial-temporal feature.

I. INTRODUCTION

The bus load forecast is a key issue for ensuring secure and reliable power systems, the goal is to predict the load demand for the next hours up to days. Most efforts in the area of electric load forecasting have been concentrated on the point forecasting by means of the previous data of each bus [1] and ignore the spatial correlations. However, the electric load influenced by grid modernization has become more active and uncertain than ever before, it is difficult to achieve a high prediction accuracy if we only consider the local bus load. For example, some external factors, such as load transfer executed by grid operators, have an influence on the spatial-temporal features of bus loads and make the loads hard to be predicted over the whole forecast horizon. Whereas it is difficult to discover the abnormal states of the power system if spatial correlations are neglected, since the load distribution of a single bus is quite regular and has a very small quantity of outliers. The spatial-temporal feature based techniques are required to recognize such abnormal operation status in the real power system and helpful to yield a high prediction accuracy. In addition, the spatial distribution of bus loads has an important influence on the stress level of power flow. The security constrained unit commitment (SCUC) also requires spatial-temporal features of bus loads to perform the determination of limits of transmission flows and evaluate security margins with possible load mismatches between demands and supplies. Due to the above reasons, the extraction of inherent
spatial-temporal patterns of bus load is required to leverage key characteristics for the single bus load prediction. In [4], a peak demand forecasting method was proposed for short-term load forecasting based on statistical methods with the inherent dynamic load characteristics. A number of algorithms such as Support Vector Regression Model [5], [6], the Random Forest Method, Kalman Filters method [7] and Artificial Neural Networks (ANNs) [8]–[10] have been employed to generate accurate prediction models associated with weather factors. In [11], the authors provided a general overview of the Recurrent Neural Networks (RNNs) employed in short term load forecasting and discussed their properties in terms of implementation and performance.

The main idea of the above methods is to use the single bus load to establish a multi-to-multi mapping network. Therefore, the spatial correlations cannot be fully represented. Reference [12] pointed out that the reliability of a system will be significantly overestimated if spatial correlations are neglected. In [13], a spatial-temporal load model was proposed for the electric power distribution facilities. In [14], based on the electric demand information of spatial correlation and temporal characteristics, a data mining technique for spatial load forecast was proposed to discover patterns in spatial data, these patterns were used later on to predict the locations of future load growth. In addition, a probabilistic risk assessment of power quality variations was proposed through the spatial and temporal behaviors of both photovoltaic generation and load demand [15].

In the meantime, little attention has been given so far to the spatial-temporal feature extraction of multiple bus loads. There are two major reasons for the lack of the research. The first is that extracting union features of multiple temporal bus loads during the data flow process is a challenging task. The second is that the description and visualization of the extracted spatial-temporal features remain problematic due to the couplings and interactions between spatial and temporal domain.

To overcome the difficulties stated above, a hybrid method composed of forecast-aided one class learning and clustering in dynamic model space is proposed in this paper. We take the spatial loads as whole and employ one-class support vector machine (SVM) [16] to recognize the change of overall operation behaviors of the power system. Dirichlet Process Mixture Model (DPMM) clustering [17] is performed to capture the spatial-temporal features of the normal load states. In general, the pattern recognition method cannot be applied directly to the feature extraction of data flow, e.g., spatial-temporal loads, hence we use the polynomial function as a surrogate model in order to represent the input data by model parameters, as it can capture the variation trend of time series load and suppress the irrelevant components such as fluctuation and noise. Then, the stable and parsimonious model parameters instead of the original load data are clustered by means of DPMM clustering in the model space. Learning in a model space helps to capture the dynamic characteristics, i.e. increasing/decreasing trend, of the bus load profiles. In addition, it has been proven that better clustering performance would be achieved in the model space than in the data space, since the data are in most cases linearly inseparable, but become sparse and separable when they are mapped into a model space [18].

The contributions of this paper can be summarized as follows:

1) A method for extracting spatial-temporal patterns of bus loads in electric grids is proposed. Then, a bus load forecast is established based on the extracted spatial-temporal features. This model can provide additional information on the variability and uncertainty associated with the future electric demand and leverage the development in respect of load forecast.

2) A state-space graph is proposed to visualize the extracted spatial-temporal patterns, which are a small number of characteristic distributions of bus loads with their evolution trends. The patterns represent all of the bus loads over the time period during which they are extracted, and also they can represent well the bus loads over a lengthy period of future time. Furthermore, the occurrences and the dominance level of each extracted pattern become very clear, the inherent physical meanings can be also illustrated.

This paper is organized as follows. Section II describes the proposed approach for abnormal load state recognition based on one-class SVM method. Section III is dedicated to the extraction of spatial-temporal features in the model space. In Section IV, a case study of the power grid in the Northeast China is conducted to illustrate the performance of the proposed method. Finally, the conclusion is presented in Section V.

II. RECOGNITION OF ABNORMAL LOAD STATES BASED ON THE ONE CLASS LEARNING

The bus loads in a real power system have very often uncertain fluctuations, volatility, outliers and high-frequency components. For example, electric power injections from renewable-based distributed generation, volatile consumptions and faulty operations would increase the probability...
of large and sudden variation in bus loads. Furthermore, the load transfer made by grid operators often happens in large cities, which brings unobvious increasing or decreasing of individual bus load. The situations above are all considered as abnormal states in the power system. But sometimes these changes might be perceived as a normal situation caused by consumption behaviors, and they are hard to be recognized if the spatial-temporal information is missing. Hence, prior to the general feature extraction, the recognition of abnormal load states should be carried out.

The spatial-temporal electric loads can be written as

\[
P = \begin{bmatrix}
p_1(t_1) & p_1(t_2) & \ldots & p_1(t_T) \\
p_2(t_1) & p_2(t_2) & \ldots & p_2(t_T) \\
\vdots & \vdots & \ddots & \vdots \\
p_N(t_1) & p_N(t_2) & \ldots & p_N(t_T)
\end{bmatrix},
\]

(1)

where \(N\) is the number of buses, \(T\) is the number of observations, and \(p_n(t_i)\) indicates the load of Bus \(n\) at time instant \(t_i\). The load state at time instant \(t_i\) is written as \(P(t_i)\), \(P(t_i) = [p_1(t_i), p_2(t_i), \ldots, p_N(t_i)]^T\). A normal load state complies with an underlying probability distribution, but an abnormal load state caused by the aforementioned situations may not share a common probability distribution with the normal states. Due to the shortage of prior information in real-time supervisory control and data acquisition (SCADA) data, the abnormal detection problem has to be formulated as an unsupervised learning problem. The main idea of the unsupervised classification scheme is to construct a hyperplane that separates the normal and abnormal states in an appropriate space. In general, this hyperplane cannot always be found in the original space, but it can be found in a feature space through a map \(\Phi: \mathbb{P} \rightarrow \mathbb{P}^*\), which maps from the original space \(P\) to a kernel space \(\mathbb{P}^*\) [19]. This feature map can be evaluated by some simple kernels, e.g., the Gaussian kernel. The strategy of one-class SVM is to map the input load data into a kernel space, and try to estimate a function \(f(\cdot)\) given by a kernel expansion in order to separate the data from the origin as far as possible. The function value is identified by evaluating which side of the hyperplane (near or far away from the origin) it falls on. A positive value indicates a normal load state, and a negative value indicates an abnormal load state.

A set of load states \(P(t_1), \ldots, P(t_T) \in \mathbb{P}\) is considered as the training data. In order to obtain the hyperplane, the following quadratic program has to be solved.

\[
\min_{\omega, \rho} \frac{1}{2} \left\| \omega \right\|^2 + \frac{1}{vT} \sum_{i=1}^{T} \xi_i - \rho,
\]

Subject to \(\omega \cdot \Phi(P(t_i)) \geq \rho - \xi_i, \quad \xi_i \geq 0, \quad i = 1, \ldots, T\).

(2)

Here, \(\omega\) is an orthogonal vector and \(\rho\) is a bias parameter of the hyperplane, \(\xi_i\) and \(v \in (0, 1)\) denote the penalty term and the upper bound on the fraction of outliers respectively. If the hyperplane with parameters \(\omega\) and \(\rho\) can be obtained by solving the optimization problem above, then the abnormal recognition can be defined by the decision function \(f(\cdot)\). For a generic test sample \(P(t)\), the decision function is formulated as

\[
f(P(t)) = \text{sgn} (\omega \cdot \Phi(P(t)) - \rho),
\]

(3)

where \(\text{sgn}(\cdot)\) is an indicator function, whereby \(\text{sgn}(x) = 1\) if \(x\) is positive, otherwise \(\text{sgn}(x) = 0\). The abnormal load states can be recognized by employing (3).

III. EXTRACTION OF SPATIAL-TEMPORAL FEATURES IN A DYNAMIC MODEL SPACE

An individual bus load has its own load characteristics following a cyclic and seasonal pattern. Although there are no definite relations among different bus loads, there exist some correlations, as stated above, due to the reasons such as common daily routines, common work schedules, or common peak load shifting. Here, the DPMM clustering is employed to capture the common characteristics of spatial load states for revealing the evolving behaviors of spatial loads. In this paper, the load data are transformed into a model space prior to cluster analysis, since it has been proven that the data are possibly more separable in the model space and better clustering performance could be achieved [18].

Polynomial function can fitting the time series of bus loads through more stable and parsimonious model parameters. The benefit of representing the original data by model parameters is that it can extract the principal features of the signal effectively and suppress the irrelevant components such as fluctuation and noise. The variation trend information represented by model parameters is taken as the cluster objectives to facilitate the dynamic feature extraction of spatial-temporal loads. By means of polynomial fitting, clustering is then implemented in the model space, instead of the original data space.

A. MAPPING INTO A MODEL SPACE

The segments of bus load data selected by a rolling window are mapped into a model space spanned by polynomials. Assuming that we are given a dataset comprising \(q\) observations of discrete points \((x_i, y_i), i = 1, 2, \ldots, q\), the data could be fitted using a polynomial function as follows

\[
y(x, \theta) = \theta_0 + \theta_1 x^1 + \cdots + \theta_M x^M = \sum_{j=0}^{M} \theta_j x^j,
\]

(4)

where \(\theta\) is the vector of model parameters, \(M\) is the order of the polynomial, and \(x^j\) denotes \(x\) raised to the power of \(j\). The model parameters can be considered as a description of the essential features of the data. A clustering process in the model space for one-dimensional data is demonstrated in Fig. 1. The signal is firstly divided into several segments, then each segment is approximated by using (4). For illustration purposes only, three selected blocks, colored by green or blue, contain a same amount of segments. The fitting model parameters associated with the two segments
can be written as a rolling window. The corresponding dense encoding format data flow are converted into a dynamic model space through of short-term bus loads. Then, the multi-dimensional load in the segment

\[ S_v = \begin{bmatrix} P_1 (t_{v-1} \times l+1) & \ldots & P_1 (t_{v} \times l) \\ P_2 (t_{v-1} \times l+1) & \ldots & P_2 (t_{v} \times l) \\ \vdots & \ddots & \vdots \\ P_N (t_{v-1} \times l+1) & \ldots & P_N (t_{v} \times l) \end{bmatrix} = \begin{bmatrix} y (t, \theta_{1,v}) \\ y (t, \theta_{2,v}) \\ \vdots \\ y (t, \theta_{N,v}) \end{bmatrix}, \]

where \( \theta_{N,v} \) is the parameter vector of the \( N \)-th bus load in the segment \( v \). Take the parameter set \( \theta_v = \{ \theta_{1,v}, \theta_{2,v}, \ldots, \theta_{N,v} \} \) as the spatial-temporal characteristics of short-term bus loads. Then, the multi-dimensional data flow are converted into a dynamic model space through a rolling window. The corresponding dense encoding format can be written as

\[ y : \mathbb{R}^{N \times l} \rightarrow \mathbb{R}^{N \times M}, \]

where \( M \ll l \). Comparing with the load state \( P (t_i) \) in the original data space, the parameter set in the model space can provide the process information during a time period. Therefore, taking the parameter set as the cluster objective facilitates the extraction of the statistical characteristics of spatial-temporal loads. The framework of clustering in a model space is illustrated in Fig. 2, where the same color blocks in a data space represent the same time period. The segments of different bus loads during the same time period

are modeled by (5) to obtain the fitting parameter sets, which are represented by the colored stars in a model space.

For an electric grid, different variation characteristics of bus loads are associated with different time periods. The main characteristics can be represented by the parameters in a model space, and the irrelevant fluctuations are suppressed.

**B. SPATIAL-TEMPORAL FEATURE EXTRACTION BASED ON THE DPMM CLUSTERING**

The DPMM is one of the most widely used statistical approaches in data clustering by introducing the Dirichlet process prior to Bayesian mixture models. That there is no need to set the number of clusters in advance gives DPMM the advantage over other cluster methods.

Let \( A \) be a sample space, and let the subsets of \( A \) satisfy \( A_1 \cup A_2 \ldots \cup A_K = A \). Assuming the observations comply with \( K \) multivariate Gaussian distributions, the likelihood parameters of the distributions are sampled from a random probability distribution \( G \), which is generated by a Dirichlet process. Then:

\[ (G (A_1), G (A_2), \ldots, G (A_K)) \sim Dir (\alpha G_0 (A_1), \alpha G_0 (A_2), \ldots, \alpha G_0 (A_K)), \]

where \( Dir (\cdot) \) denotes the Dirichlet probability density function, \( G_0 \) and \( \alpha \) are the base distribution and concentration parameter of Dirichlet probability respectively, \( \alpha > 0 \). For a given observation data set \( \{x_i\}_{i=1}^{L} \), \( L \) is the number of observations, the DPMM can be described as follows:

\[ x_i \sim F (x \mid \varphi_i), \quad \varphi_i \sim G, \quad G \sim Dir (\alpha, G_0), \]

where \( F (x \mid \varphi_i) \) is a joint distribution, \( \varphi_i \) is the parameter of the distribution. By integrating \( G \), the joint distribution of the mixture component parameters reveals a clustering effect. Assuming that the data could be classified into \( K \) specific categories, each cluster corresponds to its own distribution with its own parameters. With the posterior values of \( \varphi_i \), the probability of each observation from
Component $c$ or Cluster $c$ can be inferred by

$$y(x, c, \varphi_c) = \sum_{k=1}^{K} G_k(\varphi_k) \prod_{i=1}^{L} F(x_i | \varphi_c) p(c_i),$$

(9)

where $c\in[1, \ldots, K]$. Given a set of parameters $(\theta_1, \ldots, \theta_v, \theta_{v+1}, \ldots)$, where each sample is a $N \times M$ dimensional vector (reshaped by $N \times M$ matrix). The DPMM clustering aims to partition them into $K$ sets, i.e. $C = \{C_1, C_2, \ldots, C_K\}$.

C. “STATE-SPACE GRAPH” BASED ON FEATURE EXPRESSION

Through the DPMM clustering in a dynamic model space, the parameters containing the similar variation characteristics will be partitioned into the same category. Each cluster reveals a distinctive pattern. Now, extract the main statistic characteristics from each cluster to exhibit its typical pattern.

First of all, the mean values of the load segments of each bus in one cluster are calculated to represent the mean load level. Given that the load will not change significantly in a short segment (half an hour for our data). Larger order of the polynomial function would cause the model overfitting. For this study, we consider the linear approximation, i.e. $M = 1$. The load trend in a cluster, reflected by the second parameter of $\theta$, could be obtained by calculating the mean value of $(\theta_1(2), \ldots, \theta_v(2), \theta_{v+1}(2), \ldots)$ of the cluster. The positive value and negative value denote a rising trend and a downtrend respectively, the magnitude of the value corresponds to “steepness degree”. The characteristic indexes of the spatial-temporal load pattern are defined as follows:

$$PM(k) = \frac{1}{N_k} \sum_{P(t) \in C_k} P(t),$$

(10)

and

$$PT(k) = \frac{1}{M_k} \sum_{\theta \in C_k} \theta(2),$$

(11)

where $PM(k)$ and $PT(k)$ denote the mean load level and mean load trend of multi-bus in Cluster $k$, both of them are $N$-dimensional vectors. $N_k$ is the number of load states belonging to Cluster $k$, and $M_k$ is the number of signal segments in Cluster $k$. In addition, the standard deviation is used to measure the uncertainty of bus loads in a cluster, indicating the level of concentration of loads in a pattern. This statistical index is defined as

$$PSD(k) = \sqrt{\frac{1}{N_k} \sum_{P(t) \in C_k} (P(t) - PM(k))^2}.$$  

(12)

These three statistical indexes constitute the basis of the state-space graph, which is used to describe the spatial-temporal feature, as shown in Fig. 4. Assuming that there are 7 buses to be taken into account, the polar diagram is divided into seven directions corresponding to the 7 buses; the red dot and yellow shadow denote the mean load level $PM(k)$ and standard deviation $PSD(k)$ of each bus load segment, respectively. The length of a green line represents load trend $PT(k)$. The green line extending to the outside circle and the inside circle denote a rising trend and a downtrend respectively, and the length of the green line corresponds to “steepness degree”.

In the state-space graph, if the bus load (e.g. Bus 3, 5, 6) has narrow yellow shadow and a short green line, it indicates that the load has a narrow fluctuation range in this pattern and be more deterministic, which can be predicted easily. If the bus load has a wide yellow shadow and a long green line, it is inferred that large standard deviation in the pattern is caused by the rapid variation ratio of loads instead of the uncertainties, such as Bus 4. The load could also be forecasted accurately. However, if the bus load has wide yellow shadow and a short green line, the distribution of the load is rather dispersed without any obvious variation trend, and there exists considerable randomness in this load during a certain period. The 7th bus load, in high load level, just belongs to this situation. Therefore, more attention should be drawn to the 7th bus load from the associated dispatching center for better generation scheduling, since the strong uncertainties of high-level load may bring high risks to the power system.

D. AN APPLICATION OF SPATIAL-TEMPORAL FEATURES TO LOAD FORECAST

The extracted typical patterns are a small number of characteristic distributions of bus loads with their evolution trends. They represent all of the bus loads over the time period during which they are extracted. Based on the mean load level $PM(k)$ and mean load trend $PT(k)$ of a pattern, the load signal model of the corresponding typical pattern can be established by a classic linear model. Through the statistical analysis of the occurrence time and occurrence frequency of each pattern, the occurrence regularity of spatial-temporal patterns can be inferred. We suppose that the bus loads over a lengthy period of future time change according to these typical patterns. The future multi-bus loads can be predicted based on the inferred occurrence regularity and the fitting load models of the patterns. The framework of the proposed methodology is illustrated in Fig. 4.
IV. TEST RESULT

To illustrate the effectiveness of our proposed method, it is tested against the real SCADA data from a power grid in the Northeast China. The time series we analyzed are collected from the transmission network of the city which is constituted of 5 districts. The data contains industrial, commercial and civil load, it follows cyclic and seasonal pattern related to daily routines and work schedules. Although there is no direct influence from one bus and another, there may existing the inter-dependencies among the multi-bus loads of a same phenomenon in different locations of the city. The data consists of 20 bus loads over 60 days with 1-min cadence, spanning from January 1, 2015 to March 1, 2015. We use the load relative to the first 50 days as training set and the load of the last 10 days as validation set. The abnormal phenomena considered in this test case is a load transfer operation, which happens in the first few days. The proposed spatial-temporal feature based load forecast method are used to predict the power load 24h ahead with 30-min cadence, which corresponds to 48 time step ahead prediction. All simulations are executed under the Python environment on a desktop PC platform equipped with Intel Core ™ i7-5500U CPU and 8G RAM.

A. RECOGNITION OF ABNORMAL STATES VIA ONE-CLASS SVM

The bus loads are resampled in a longer interval in order to prepare for the polynomial fitting. The segment length of load curves was chosen according to the physical properties of power loads. The load time series generated by the dynamics of the system can vary significantly and be unique during a long observation period, therefore, selecting long segment length is not good for the common characteristic extraction. Whereas it’s stationary without obvious variation in a very small time interval. We select 30-min cadence as the segment length in the test and certainly less or more than 30-min cadence will also be accepted. Then, the data of the first 50 days are resampled in 30-min cadence and are normalized as shown in Fig. 5.

Fig. 5 shows that the bus loads follow daily cycle and weekly cycle except for the first few days. The abnormal fluctuations in the first few days were caused by a load transfer operation. The load states before and after the load transfer operation are rather different, and thus the load spatial-temporal features could not be extracted from the raw data directly. For the improved and generalized feature extraction, one-class SVM is adopted here to identify abnormal load states.

In our case, it implies that the radial basis function (RBF) kernel was selected in one-class SVM and the parameter $\nu$ is chosen as 0.05. The parameter $\nu$ refers to the upper bound on the fraction of outliers. More experiments have been performed to observe the effect of using a different value of $\nu$ on the recognition performance, $\nu$ is selected between 0.01 and 0.2 with 0.01 step. Due to the space limitation, detailed results of the test are not presented. According to the experiments, it is observed that the recognition result does not have a meaningful relation with the increasing value of $\nu$, when it is larger than 0.05. Since there are a small quantity of abnormal load states of the power grid, even if the upper bound on the fraction of outliers is set to a higher value, the results will not change considerably. However, the performance of one-class SVM is degraded, when a smaller $\nu$ (lower than 0.03) is selected, since some abnormal states are mistaken for the normal ones. Therefore, $\nu$ is set as 0.05, which is also considered by the classic literature [20].

Fig. 6 shows the recognition result of the abnormal load states, and ‘1’ and ‘0’ indicate abnormal state and normal state.
state respectively. It is observed that the recognized abnormal states mainly concentrate in the first few days, the results of one-class SVM method are in agreement with actual fact by contrasting the start-stop time of real event and the start-stop time of the dense recognized abnormal states. In addition, it is noted that a few identified abnormal states appear after the load transfer operation, their dispersed distribution implies that the unusual increments (or decrements) of loads are probably related to the consumption behaviors and not caused by specific external events. Removing these abnormal states is also good for further spatial-temporal feature extraction.

B. EXTRACTION OF THE SPATIAL-TEMPORAL FEATURES

The data from the 10th day to the 50th day are applied for the extraction of spatial-temporal characteristics. Because of the load transfer situation, there is not any treatment for the abnormal state in the first 10 days. We switch back to 1-min cadence instead of 30-min in order to fit the models through the identified normal load states. The load fluctuation features of the normal ones are learned in a dynamic model space implemented by polynomial fitting through a series of signal segments of multiple bus loads selected by a rolling window (30-min). Afterwards, the parameters of different segments are employed for the DPMM clustering.

The spatial-temporal features extracted from the dynamic model space are visualized by the state-space graphs shown in Fig. 7, where five different colors represent five different geographical districts in the city. They exhibit 9 operation patterns containing the load averages, load variation trends and load uncertainty. Fig. 8 shows when the 9 patterns occur within the 24 hours of each day from the 10th day to the 50th day, where the color bar indicates the number of clusters, “0” represents the abnormal state. The patterns with daily cycle and weekly cycle can be observed.

From Fig. 7 and Fig. 8, it is noted that Pattern 1 always occurs in the early morning (about 0:00am to 5:00am). In this pattern, all of the bus loads except for three bus loads are at a low load level and low level of fluctuation. Pattern 2 shows that the three bus loads start to drop and the others increase slightly from 5:00am to 6:30am. Then, the power system turns to Pattern 3 and lasts for one hour. Pattern 3 shows that all the bus loads are at a medium level with rapid variation, which are inferred by the long green lines (calculated by PT(k)). It’s worth noting that Pattern 5 is the only pattern that appears twice on weekdays but disappears on weekends. According to Pattern 5, the daily peak load of most buses occurs from about 8:30am to 12:00am and from 13:00pm to 16:30pm with low volatility and weak trend judging from the narrow yellow shadows and short green lines. It is mainly related to some industrial consumption during daytime. For Pattern 6, the three bus loads in the bottom left of the state-space graph are inferred to have high uncertainty and would be difficult to forecast. As can be seen from Fig. 7, Pattern 8 and Pattern 3 share similar average loads, but the trends of their loads are different. From 22:00 to 24:00, Pattern 9 with mid-night valley appears. Most of the bus loads are in the fast transition from early night hours to valley hours mainly due to residential loads supplied by the associated substations. After that, it turns back to Pattern 1 starting another day.

Fig. 8 shows the occurrence regularity of the extracted patterns on weekdays and weekends. For weekdays, with each half hour as one time slot there are 48 time slots in a day. Starting from 0:00 to end at 24:00 the time when each pattern occurs follows this sequence: Pattern 1 occurs from 00:00 to 5:00; Pattern 2 occurs from 5:00 to 6:30; Pattern 3 occurs from 6:30 to 7:30; Pattern 4 occurs from 7:30 to 8:30; Pattern 5 occurs from 8:30 to 12:00; Pattern 6 occurs from 12:00 to 13:00; Pattern 5 occurs again from 13:00 to 16:30; Pattern 7 occurs from 16:30 to 20:30; Pattern 8 occurs from 20:30 to 22:00; Pattern 9 occurs from 22:00 to 24:00.

C. LOAD FORECAST BASED ON THE SPATIAL-TEMPORAL FEATURES

To illustrate the degree of the representativeness of the spatial-temporal patterns, the extracted patterns are used to forecast the multi-bus loads 24h ahead for the next 10 days. The proposed hybrid method is established based on one-class SVM and clustering in the model space, named as one-class SVM – Model Space (OCS-MS). Here SVM, Neural Network (NN) and Long-Short Term Memory (LSTM) are employed as a benchmark. The SVM and NN are popular load forecast methods due to their flexibility as nonlinear predictors. LSTM is the advanced class of the state-of-the-art Recurrent Neural Networks. The proposed hybrid method has been examined on 20 buses. The training period is the first 50 days for all methods. Thus, each forecast method has 50 × 48 = 2400 training samples and is executed for 24h ahead prediction. Bus loads are updated once every 24 hours, and the parameters of these models are re-estimated every day.

The mean absolute percentage error (MAPE) is used as the evaluation criteria for measuring the performance of load forecast. It is defined as,

$$\text{MAPE} = \frac{1}{N} \sum_{i=1}^{N} \frac{|\hat{x}(i) - x(i)|}{x(i)},$$  

(13)
where $\hat{x}(i)$ and $x(i)$ denote the predicted load and actual load; $i$ is the sample number, and $N$ is the total number of the samples.

Fig. 9 shows the load forecast performance of different methods. As seen from the MAPE values, it is observed that the proposed OCS-MS outperforms the SVM model for any of the 20 buses and achieves a higher prediction accuracy than the NN method for the most buses. One of the reasons that the proposed method has considerable improvement compared with SVM and NN is that spatial abnormal load state recognition has been considered. The existence of abnormal load distributions, which does not comply with the probability distribution of the normal ones, is unfavorable for general feature extraction. Sometimes bus load forecasting encounters high prediction errors due to the run-state change of spatial loads, which is hard to be dealt with by the single bus based load forecasting method. Therefore the one-class SVM strategy is necessary, especially when the prior information of the measured data is not available. Even though there are no abnormal load states, the one-class SVM strategy cannot have a negative effect on the further treatment.

In Fig. 9, it is noted that the proposed method has a higher forecast accuracy with respect to the LSTM for some buses, but be poorer than it for some other buses. The average MAPE of the OCS-MS and the average MAPE of the LSTM for the 20 buses are 0.058 and 0.057 respectively. They achieve very similar forecast accuracy on this dataset. The LSTM method is the state-of-the-art technique in time series prediction problems, although the prediction accuracy of the proposed method is not obviously improved, it still shows satisfactory predictive ability. Moreover, the proposed method achieves benefits in other aspects through comparison against the time series based benchmarks.

Firstly, the outputs of the proposed method are the spatial-temporal distributions with information on the variation trend and uncertainty associated with the future electric demand, instead of the only temporal load values. It is a much more
complex task and be crucial to the SCUC and the dispatching centers of power systems. The index of uncertainty extracted by the proposed method can provide additional information on the reliability of forecast result, which plays an important role in forecast performance.

Secondly, the point forecasting techniques based on the artificial intelligence, e.g., NN and LSTM, focus on obtaining accurate expected values and serve as a black box; the structure and parameters of models are the major factors affecting the prediction accuracy and hard to be explained physically. The proposed method provides the spatial-temporal features, which are a small number of characteristic distributions of bus loads with their evolution trends and can be visualized. They reveal the changing process of spatial bus loads and have inherent physical meanings, and also provide a better understanding of the dynamics of the monitored system.

V. CONCLUSION
In this paper, we propose a spatial-temporal feature based bus load forecasting methodology. To address the challenge of coupling and interactions in the multi-bus loads between time domain and space domain, one-class SVM was adopted to recognize the spatial abnormal load states, and the DPMM cluster method in a dynamic model space was proposed. According to our experimental results, it shows that there exist correlations among the spatial load distribution, which are expressed by the proposed state-space graphs. The extracted state-space graphs reveal the changing process of spatial bus loads and have clear physical meanings. Through the spatial correlation and temporal load characteristics analysis, the prediction accuracy of load forecast has been improved. The method is tested against a set of real data from a power grid of the Northeast China. The testing results showed that the proposed bus load forecast method gave clear temporal trends and spatial distributions of multi-bus loads and achieved satisfactory accuracy. The practical value of the hybrid method is significant in the sense that it can leverage the development of time series based load forecasting.

REFERENCES
[1] I. Moghram and S. Rahman, “Analysis and evaluation of five short-term load forecasting techniques,” IEEE Trans. Power Syst., vol. 4, no. 4, pp. 1484–1491, Nov. 1999.
[2] Z. Xie, G. Manimaran, V. Vital, A. Phadke, and V. Centeno, “An information architecture for future power systems and its reliability analysis,” IEEE Trans. Power Syst., vol. 17, no. 3, pp. 857–863, Aug. 2002.
[3] Y. Fu, M. Shahidehpour, and Z. Li, “Security–constrained unit commitment with AC constraints,” IEEE Trans. Power Syst., vol. 20, no. 2, pp. 1001–1013, May 2005.
[4] E. Barakat, M. Qayyum, M. Hamed, and S. Al Rashed, “Short-term peak demand forecasting in fast developing utility with inherit dynamic load characteristics. I. Application of classical time-series methods. II. Improved modelling of system dynamic load characteristics,” IEEE Trans. Power Syst., vol. 5, no. 3, pp. 813–824, Aug. 1990.
[5] E. E. Elattar, J. Goulermas, and Q. H. Wu, “Electric load forecasting based on locally weighted support vector regression,” IEEE Trans. Syst., Man, Cybern. C, Appl. Rev., vol. 40, no. 4, pp. 438–447, Jul. 2010.
[6] H. Jiang, Y. Zhang, E. Muljadi, J. J. Zhang, and D. W. Gao, “A short-term and high-resolution distribution system load forecasting approach using support vector regression with hybrid parameters optimization,” IEEE Trans. Smart Grid, vol. 9, no. 4, pp. 3341–3350, Jul. 2018.
[7] C. Guan, P. B. Luh, L. D. Michel, and Z. Chi, “Hybrid Kalman filters for very short-term load forecasting and prediction interval estimation,” IEEE Trans. Power Syst., vol. 28, no. 4, pp. 3806–3817, Nov. 2013.
[8] J. Jiménez, K. Donado, and C. G. Quintero, “A methodology for short-term load forecasting,” IEEE Latin Amer. Trans., vol. 15, no. 3, pp. 400–407, Mar. 2017.
[9] P. Siano, C. Cecati, H. Yu, and J. Kolbusz, “Real time operation of smart grids via FCN networks and optimal power flow,” IEEE Trans. Ind. Inf., vol. 8, no. 4, pp. 944–952, Nov. 2012.
[10] C. Cecati, J. Kolbusz, P. Rozycki, P. Siano, and B. M. Wilamowski, “A novel RBF training algorithm for short–term electric load forecasting and comparative studies,” IEEE Trans. Ind. Electron., vol. 62, no. 10, pp. 6519–6529, Oct. 2015.
[11] F. M. Bianchi, E. Maiorino, M. C. Kampffmeyer, A. Rizzi, and R. Jenssen, “An overview and comparative analysis of recurrent neural networks for short term load forecasting,” CoRR, vol. abs/1705.04378, May 2017. [Online]. Available: http://arxiv.org/abs/1705.04378
[12] B. Seguinot and A. Zani, “The role of spatial correlations in Monte Carlo studies on power system,” in Proc. IEEE Eindhoven PowerTech, Jun. 2015, pp. 1–6.
[13] J.-H. Shin, B.-J. Yi, Y.-I. Kim, H.-G. Lee, and K. H. Ryu, “Spatiotemporal load–analysis model for electric power distribution facilities using consumer meter–reading data,” IEEE Trans. Power Del., vol. 26, no. 2, pp. 736–743, Apr. 2011.
[14] H.-C. Wu and C.-N. Lu, “A data mining approach for spatial modeling in small area load forecast,” IEEE Trans. Power Syst., vol. 17, no. 2, pp. 516–521, May 2002.
[15] S. Pukhrem, M. Basu, and M. F. Conlon, “Probabilistic risk assessment of power quality variations and events under temporal and spatial characteristic of increased PV integration in low–voltage distribution networks,” IEEE Trans. Power Syst., vol. 33, no. 3, pp. 3246–3254, May 2018.
[16] G. Ratsch, S. Mika, B. Scholkopf, and K.-R. Muller, “Constructing boosting algorithms from SVMs: An application to one-class classification,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 24, no. 9, pp. 1184–1199, Sep. 2002.
[17] N. Bouguila and D. Ziou, “A Dirichlet process mixture of generalized Dirichlet distributions for proportional data modeling,” IEEE Trans. Neural Netw., vol. 21, no. 1, pp. 107–122, Jan. 2010.
[18] A. Bosch, A. Zisserman, and X. Muoz, “Scene classification using a hybrid generative/discriminative approach,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 30, no. 4, pp. 712–727, Apr. 2008.
[19] R. Perdisci, G. Gu, and W. Lee, “Using an ensemble of one-class SVM classifiers to harden payload-based anomaly detection systems,” in Proc. 6th Int. Conf. Data Mining (ICDM), Hong Kong, Dec. 2006, pp. 488–498.
[20] B. Scholkopf, R. C. Williamson, A. J. Smola, J. Shawe-Taylor, and J. C. Platt, “Support vector method for novelty detection,” in Advances in Neural Information Processing Systems, S. A. Solla, T. K. Leen, and K.-R. Muller, Eds. Cambridge, MA, USA: MIT Press, 2000, pp. 582–588.

WEI ZHANG received the B.S. degree in signal and information processing from the University of Electronic Science and Technology of China, in 2013. She is currently pursuing the Ph.D. degree in electrical engineering from Northeast Electric Power University, China. She is also a Teacher with Northeast Electric Power University. Her research interests include spatial-temporal features of electric load, load forecasting, and deep learning applications to power systems.
GANG MU received the Ph.D. degree in electrical engineering from Tsinghua University, China, in 1991. He is currently a Professor with Northeast Electric Power University, China. His research interests include power system stability analysis, planning and operation analysis of large-scale wind farms integrating to power systems, planning and control of energy storage, and big data/AI applications in power system operation. He is a Fellow and a Councilor of CSEE. He is also the Director of the Modern Power System Simulation and Control and Renewable Energy Technology Key Laboratory of Ministry of Education China.

CHEN SONG received the B.S. degree in biomechanics from Claude Bernard University Lyon 1, France, and the Ph.D. degree from Heidelberg University, Germany, in 2018. He is currently a postdoctoral position at the Engineering Mathematics and Computing Lab (EMCL), Interdisciplinary Center for Scientific Computing (IWR), Heidelberg University, and the Heidelberg Institute for Theoretical Studies, Heidelberg, Germany. His research interests include computational fluid dynamics, uncertainty quantification, and variational multiscale method.

GANGUI YAN received the Ph.D. degree in electrical engineering from Tsinghua University, China, in 2003. He is currently a Professor with Northeast Electric Power University, China. His research interests include the wind power generation and power electronic technology.

VINCENT HEUVELINE received the Ph.D. degree in computer science from the Université de Rennes and INRIA (Institut National de Recherche en Informatique et Automatique), in 1997, and the Habilitation degree in mathematics from Heidelberg University, in 2002. Since 2004, he has been a Professor with Karlsruhe University and KIT, until he moved to Heidelberg, in May 2013. Besides his professorship at IWR, he is currently the Chief Information Officer (CIO) and the Director of the Engineering Mathematics and Computing Lab (EMCL), Interdisciplinary Center for Scientific Computing (IWR), Heidelberg University. He is also the Leader of the Research Group Data Mining and Uncertainty Quantification (DMQ) at the Heidelberg Institute for Theoretical Studies.