A Short-Term and High-Resolution Distribution System Load Forecasting Approach Using Support Vector Regression with Hybrid Parameters Optimization

[1]Huaiguang Jiang, *Student Member*, IEEE, [1]Yingchen Zhang, *Senior Member*, IEEE, [1]Eduard Muljadi, *Fellow*, IEEE, [2]Jun Jason Zhang, *Senior Member*, IEEE, [2]David Wenzhong Gao, *Senior Member*, IEEE

***Abstract*—This paper proposes an approach for distribution system load forecasting, which aims to provide highly accurate short-term load forecasting with high resolution utilizing a support vector regression (SVR) based forecaster and a two-step hybrid parameters optimization method. Specifically, because the load profiles in distribution systems contain abrupt deviations, a data normalization is designed as the pretreatment for the collected historical load data. Then an SVR model is trained by the load data to forecast the future load. For better performance of SVR, a two-step hybrid optimization algorithm is proposed to determine the best parameters. In the first step of the hybrid optimization algorithm, a designed grid traverse algorithm (GTA) is used to narrow the parameters searching area from a global to local space. In the second step, based on the result of the GTA, particle swarm optimization (PSO) is used to determine the best parameters in the local parameter space. After the best parameters are determined, the SVR model is used to forecast the short-term load deviation in the distribution system. The performance of the proposed approach is compared to some classic methods in later sections of the paper.**

***Index terms*—** Short-term load forecast, support vector regression, grid traverse algorithm, particle swarm optimization, distribution system

List of Symbols

$\mathbf{L}$	The historical load data vector of a distribution system, where L_i is the ith element of set $\mathbf{L}$, $i \in \{1, \cdots, n\}$.
$\mathbf{L}'$	The normalized historical load data vector of a distribution system, where L'_i is the ith element of set $\mathbf{L}'$, $i \in \{1, \cdots, n\}$.
$\hat{\mathbf{L}}$	The forecasting load data vector of a distribution system, where $\hat{L}_{i_1}$ is the i_1th element of set $\hat{\mathbf{L}}$, $i_1 \in \{n+1, \cdots, n+m\}$.
$L_{\max}$, $L_{\min}$	The maximum value and minimum value of $\mathbf{L}$, respectively.
$L'_{\max}$, $L'_{\min}$	The maximum value and minimum value of $\mathbf{L}'$, respectively.
L'_{i_2}	L'_{i_2} is the i_2th element of training load data, $L'_{i_2} \in \mathbf{L}'$, i_2 is a time index of training data, $i_2 \in \{1, \cdots, n\}$
K_{pre}	A normalization mapping ratio for data pretreatment.
$\Phi(\cdot)$	A nonlinear mapping function of SVR, which is defined to map the input data from space $R^{n'}$ to space $R^{m'}$.
$\mathbf{R}^{n'}$	A real coordinate space with dimension n'.
	A parameter vector of SVR, where $_j$ is the jth element of , $j \in \{1, \cdots, p\}$.
$\bar{}$, $\underline{}$	The upper limit vector and lower limit vector of .
$\boldsymbol{\Delta}$	A searching step vector for GTA, where Δ_j is the jth element of $\boldsymbol{\Delta}$.
$\boldsymbol{\Lambda}$	A high-dimensional vector for GTA, where Λ_j is the jth element of $\boldsymbol{\Lambda}$.
$\mathbf{H}$	A traversing vector, which contains all the combinations of different parameters in $\boldsymbol{\Lambda}$.
m_j	It is defined that the total number of the elements in vector Λ_j is m_j.
H_{j_2}	H_{j_2} is the j_2th element of $\mathbf{H}$, where $j_2 \in \{1, 2, \cdots, m_1 \times m_2 \times, \cdots, \times m_p\}$.
β_{j,m''_j}	The parameter β_{j,m''_j} is the m''_jth element chosen from vector Λ_j, where $m''_j \in \{1, 2, \cdots, m_j\}$.
H_b	The best parameter vector for SVR, H_b is an element in $\mathbf{H}$.
ω	ω is a linear combination coefficient of SVR, $\frac{1}{2}\omega^T\omega$ indicates the flatness of the regression coefficients.
b	b is an offset coefficient of SVR.
C, γ, ε	C is a trade-off parameter, γ is a parameter of (Gaussian) radial basis function, and ε is an adjustable precision parameter, which indicates the training error threshold.
R_{CV}	The objective function of designed Cross-Validation.
$\hat{\zeta}_\tau$	$\hat{\zeta}_\tau$ is the load data forecasted by proposed approach to test in the Cross-Validation, and ζ_τ is the corresponding load data.
$\boldsymbol{\alpha}_{i_4}, \boldsymbol{\nu}_{i_4}, \boldsymbol{\eta}_{i_4}$	The vector of position, the vector of velocity, and the vector of the best historical position of the i_4th particle in PSO, respectively,
φ_1, φ_2	Two positive acceleration coefficients for velocity updates of each particle in PSO.
θ_1, θ_2	Two independently random variables with uniformly distributed range $(0, 1)$ for velocity updates of each particle in PSO.
η_g	The best position vector among all particles in PSO.
S_{skw}, S_{krt}	The skewness and kurtosis of the proposed approach, respectively.

I. Introduction

As forms of energy and power systems evolve, new distribution systems in the future, featuring distributed renewable energy generation and demand-response control, will be able to operate independently from the bulk power system [1]–[5]. The load profiles of the distribution systems will contain more stochastically abrupt deviations because the behavior of

H. Jiang, Y. Zhang (Corresponding author), and E. Muljadi are with the National Renewable Energy Laboratory, Golden, CO 80401 USA, (e-mail:yingchen.zhang@nrel.gov). J. Zhang and W. Gao are with the University of Denver, CO 80210 USA.

the end users has much bigger impact on distribution systems than it does on transmission systems. Successfully operating an independent distribution system requires much accurate and high-resolution load forecasting than today's technique can provide [6]–[12]. In this paper, a short-term and high-resolution load forecasting is proposed for the distribution feeders, which contains the aggregated loads of a small section of distribution feeder load.

There are well established methods for load forecasting in general. In [13], an algorithm based on support vector machine (SVM) with simulated annealing is employed to forecast the system load. In [14], the ant colony optimization is used to determine the best parameters for the SVR, which is implemented to forecast the system load. A load forecasting algorithm based on SVM and genetic algorithm (GA) is presented in [15]. Short-term load forecasting normally has more computational intensive solutions. For example, the ANN-based short-term load forecast approaches are presented in [16], [17]. In [18], an autoregressive integrated moving-average (ARIMA) model is used to forecast the short-term load. A Kalman filtering-based short-term load forecasting approach is presented in [19]. In [20], an approach based on double seasonal exponential smoothing is used for short-term load forecasting.

Although many load forecasting approaches exist, most of them focus on load forecasting in transmission systems. In transmission systems, the aggregated loads are three-phase balanced and always fairly smooth.In this paper, a small section of distribution feeder load contains several end customers, which are three-phase unbalanced and much less than the aggregated loads at a transmission feeder. This indicates that the loads profile characteristics of a distribution feeder are different. In [21], [22], the impact of electric vehicles on the distribution systems is evaluated, and it indicates that abruptly stochastic load deviation is a feature of distribution systems. In addition, combined with distributed renewable energy resources, the control strategy of the distribution system requires a high-speed and high-resolution approach for stochastical nonlinear tracking and forecasting [6], [7], [23], [24]. Therefore, it is critical and imperative to provide a short-term and high-resolution forecasting approach for the aggregated loads of a small section of distribution feeder load.

In this paper, the proposed SVR-based short-term load forecasting approach is a supervised machine learning approach that requires the preprocessing of the input data to achieve better regression and forecasting performance. After preprocessing, the collected historical load data are used to train the SVR model. Meanwhile, compared to GA, ANN, and simulated annealing approach, particle swarm optimization (PSO) has a higher converge speed to determine the best parameters for SVR, but the PSO cannot guarantee to converge to a global optimal solution [25], [26]. To solve the above problems, a two-step hybrid parameters searching algorithm is proposed to determine the parameters effectively and efficiently.

The paper is organized as follows. In Section II, the problem of short-term distribution system load forecasting is formulated and the flowchart of the proposed approach is introduced. In Section III, the designed SVR-based load forecasting is introduced. In Section IV, the two-step hybrid parameter optimization method is illustrated for SVR-based forecaster. In Section V, numerical results of the proposed approach are presented and compared to other methods.

II. Problem Formulation

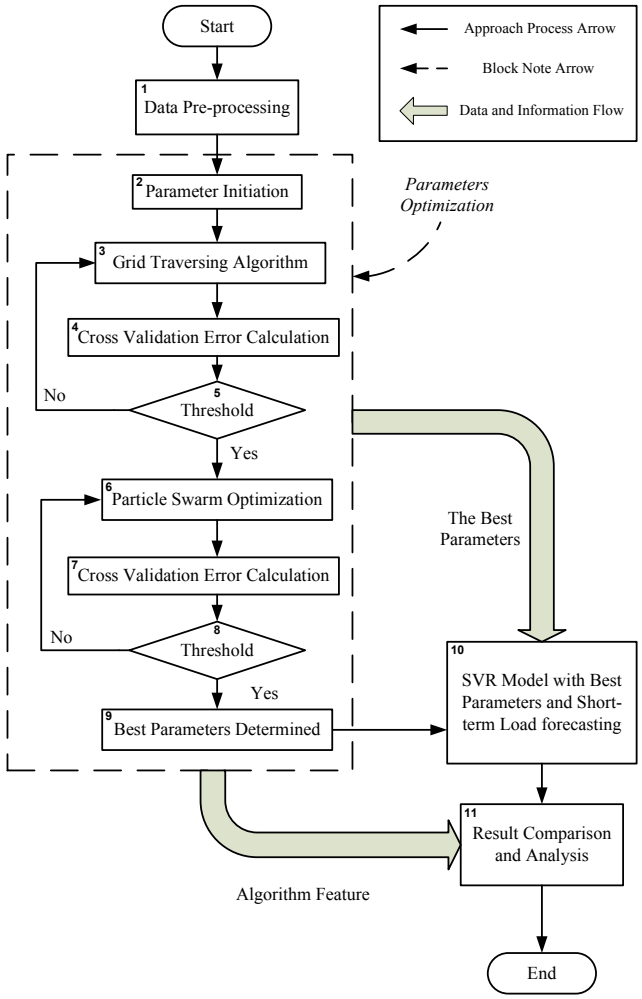


Fig. 1. Flowchart of the proposed short-term load forecasting approach.

In this paper, the proposed load forecasting approach consists of three major parts: data preprocessing, SVR model training, and forecasting with the SVR model. To determine the best parameters, the proposed two-step hybrid parameter optimization method is designed in the SVR model training part.

The flowchart of the proposed approach is shown in Fig. 1. Firstly, the collected historical load data is processed to reduce the excessive deviation. Next, a two-step approach is utilized to optimize parameters for the SVR solution. In the first step, the grid traverse algorithm (GTA) traverses the global solution space into local spaces, which can be process in parallel and narrows the parameters searching area. In the second step, the PSO is used to determine the best parameters of the SVR in the local solution space. If multiple local solution spaces

are determined by the GTA, the PSO can calculate the best parameters in the local solution spaces in parallel. After the two-step parameter optimization, a best parameter vector can be decided for the SVR. Finally the future load is forecasted based on the determined SVR function and historical data. The detail procedures of the proposed approach can be found in Appendix A.

III. SVR for Short-term Distribution System Load Forecasting

A. Data preprocessing

The original data from the natural world usually contains a lot of excessively deviating instances, which distract the target of the supervised forecasting approach and significantly impact performance [27]. In this paper, we are trying to forecast load at feeder level or distribution system level, therefore it is necessary to process the original load data to improve the performance of the proposed approach. The detail procedures can be found in Appendix B.

B. Basic Concept of SVR

Based on support vector machine, the SVR is a supervised learning method used for nonlinear regression and time series forecasting applications [28]–[30]. The historical data is treated as input training data, and the forecasting is based on the SVR model trained by historical data.

C. SVR for Short-Term Load Forecasting in Distribution Systems

In a SVR model, a nonlinear mapping function $\Phi(\cdot)$: $R^{n'} \rightarrow R^{m'}$ is defined to map the input data into a high-dimensional feature space, where $n', m' \in \{1, 2, 3, \cdots\}$, $m' > n'$. In [30], [31], it is noted that the (Gaussian) radial basis function (RBF) kernel is widespread with high performance for regression and forecasting. In this paper, the RBF is designed as:

$$\begin{aligned} K_{RBF}(x_1, x_2) &= \Phi(x_1)^T \Phi(x_2) \\ &= \exp(-\gamma ||x_1 - x_2||^2) \end{aligned} \tag{1}$$

where the x_1 and x_2 are the inputs, in this paper, they are the historical load data, and $\gamma \in R$ is an adjustable parameter.

In this paper, the training data is the historical load data $\mathbf{L}$ in time series; after preprocessing, the training data set becomes $\mathbf{L}'$ in time series. The goal of the SVR is to find a function f, which can be formulated as follows:

$$f(x) = \langle \omega, \Phi(x) \rangle + b, \tag{2}$$

where $f(x)$ is the regression function aiming to have at most ε deviation from the elements of the collected historical load data $\mathbf{L}'$, ω is a linear combination coefficient, b is an offset coefficient, ε is an adjustable precision parameter, which indicates the training threshold.

To improve the robustness of the algorithm, the slack variables ξ_{i_2} and $\xi^*_{i_2}$ are introduced as in [28]. The risk function of the SVR can be formulated as follows:

$$R_{risk} = \min_{\varepsilon, \omega, \xi_{i_2}, \xi^*_{i_2}, C, b, \gamma} \{ \frac{1}{2} \omega^T \omega + C \sum_{i_2=1}^{n} (\xi_{i_2} + \xi^*_{i_2}) \} \tag{3}$$

Subject to

$$\begin{cases} L'_{i_2} - f(x_{i_2}) \leqslant \varepsilon + \xi_{i_2}, \\ -L'_{i_2} + f(x_{i_2}) \leqslant \varepsilon + \xi^*_{i_2}, \\ \xi_{i_2}, \xi^*_{i_2} \geqslant 0. \end{cases} \tag{4}$$

Here in (3), i_2 is a time index of training data, $i_2 \in \{1, \cdots, n\}$, $\frac{1}{2}\omega^T\omega$ indicates the flatness of the regression coefficients, the second item is the cost function of the training errors between L'_{i_2} and $f(x_{i_2})$, C is a trade-off coefficient between the first two items, ξ_{i_2} and $\xi^*_{i_2}$ indicates the training error above the precision parameter ε and below $-\varepsilon$, respectively. In [29], the risk function (3) with the constraint (4) can be solved as a dual problem, which can minimize the risk function and determine ω, b, ξ_{i_2}, and $\xi^*_{i_2}$. Therefore, the parameters γ, C, and ε are chosen as the optimal parameters, which are most important to the accuracy of the forecasting [15], [28], [30], [31]. The following chapter introduces a two-step hybrid method for parameters optimization.

IV. Parameter Optimization of SVR

Compared to the GA, ANN, and simulated annealing approaches in [13], [15], [16], [32], PSO is an effective algorithm for solving nonlinear problems; however, it does not guarantee that the algorithm can converge to a global optimal solution in the solution space [25]. In [26], it is noticed that the conventional PSO may become divergent or trapped in a local optimal solution. Conventionally, decreasing the global search step size and increasing the number of particles are two solutions to avoid local optimizations, but they may result in prohibit computational load. Therefore, considering both the global convergence speed and accuracy of the algorithm, a two-step hybrid global optimization algorithm is proposed to determine the best forecasting parameters for the SVR.

A. GTA for Parameter Optimization

1) Design of GTA: In GTA, based on the parameter vector , its upper and lower limit vectors ¯, _, and the searching step vector $\mathbf{\Delta}$, a traversing high-dimensional vector $\mathbf{\Lambda}$ and a traversing vector $\mathbf{H}$ can be built as follows.

At the beginning, for each parameter $_j$ in the parameter vector , a vector Λ_j is built with the corresponding step Δ_j as follows

$$\Lambda_j = [\underline{\Omega}_j \ \cdots \ \__j + (m''_j - 1)\frac{(\bar{}_j - \__j)}{\Delta_j} \ \cdots \ \bar{}_j] \tag{5}$$

where

- j is the index of parameters in , and $j \in \{1, 2, \cdots, p\}$.
- Λ_j is a vector corresponding to parameter $_j$, and computed with its upper and lower limits $\__j$, $\bar{}_j$, and searching step Δ_j.
- It is defined that the total number of the elements in vector Λ_j is m_j, and $\frac{(\bar{}_j - \__j)}{\Delta_j}$ is designed as an integer.
- $\__j + (m''_j - 1)\frac{(\bar{}_j - \__j)}{\Delta_j}$ is the m''_jth element in Λ_j, where $m''_j \in \{1, 2, \cdots, m_j\}$

After all the parameter vectors Λ_j, $j \in \{1, 2, \cdots, p\}$ are built with (5), a high-dimensional vector $\mathbf{\Lambda} \in$

$R^{m_1 \times m_2 \times, \cdots, \times m_p}$ is built. Then, a traversing vector $\mathbf{H}$ is built as a finite multi-Cartesian product, which is illustrated in Appendix C.

2) Design of Cross-Validation: A designed cross-validation is used with the GTA to limit the overfitting problem and compute the best parameters [33]. The training data is divided into 10 groups; for each time, one group is selected as the test data, and the rest of the data are training data. The objective function is designed as a prediction sum of squares (PRESS) statistic

$$R_{CV} = \min_{\gamma,\varepsilon,C} \sum_{\tau=2}^{10} \frac{(\zeta_\tau - \hat{\zeta}_\tau)^2}{10} \tag{6}$$

where $\hat{\zeta}_\tau$ is the load data forecasted by proposed approach to test in the Cross-Validation, and ζ_τ is the corresponding load data. For example, if $\tau = 2$, this means the data group 1 is the training data to forecast data group 2; if $\tau = 5$, the data groups $1, 2, 3, 4$ are training data to forecast data group 5. The difference between ζ_τ and $\hat{\zeta}_\tau$ is the residual error. In addition, because the normalized load data ranged at $(0, 1)$ in the preprocessing block, this objective function has the same meaning as the mean square percentage error (MSPE), which can indicate the accuracy and deviation of the forecasting results [34], [35]. The detailed description of the GTA can be found in Appendix D.

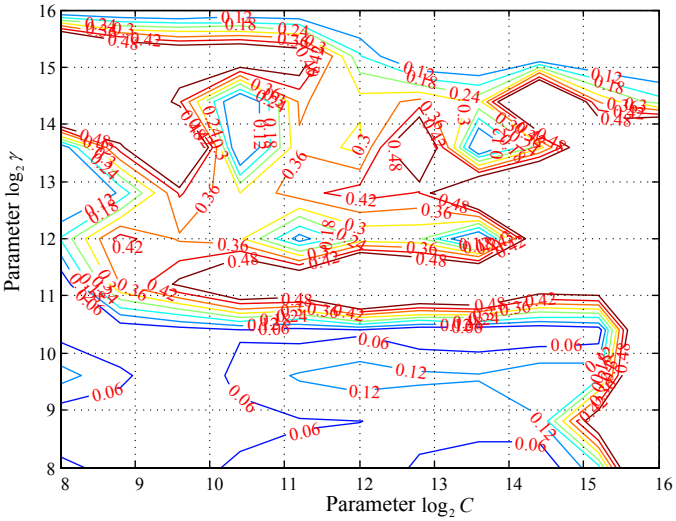


Fig. 2. A contour map of parameter selection of the GTA for a distribution system load data prediction (ε = 0.001).

3) Parameter Selection with GTA: A Distribution System Load Data Prediction: A set of load data of a distribution system is employed as the training and testing data. The sample rate is 1 Hz, the time period is 1 hour, and the data length is $3,600$. 85% of the data is used to train the SVR model, and 15% of the data is used to test the prediction results. According to the discussion above, the parameter vector = $[\gamma\ C\ \varepsilon]$, and in this scenario ε is set as 0.001. The upper limit vector ¯ = $[2^{16}\ 2^{16}]$, the lower limit vector _ = $[2^8\ 2^8]$, and the searching step vector $\mathbf{\Delta}$ = $[2\ 2]$. The traversing grid vector $\mathbf{H}$ can be built with this information. The results of the parameter selection with the GTA are shown as in Fig. 2.

In Fig. 2, the x-axis denotes the parameter $\log_2 C$, the y-axis indicates the parameter $\log_2 \gamma$, and the curves with red to blue indicate the PRESS value from large to small. It is clear that the areas with blue curves have smaller PRESS values and the next step optimization will only process this local area. In this case, the lower limit and upper limit of C are 2^8 and 2^{15}, respectively. The lower limit and upper limit of γ are 2^8 and 2^{11}, respectively.

B. PSO for the Optimal Parameter Selection

1) Basic Concept of the PSO: PSO is a numerical intelligence-based approach, which can solve many nonlinear optimal problems that cannot be solved analytically. In addition, PSO has a higher convergency speed in the solution space, convenient to implement, and its converge ability is not largely affected by the size and nonlinearity of the optimal problems [36].

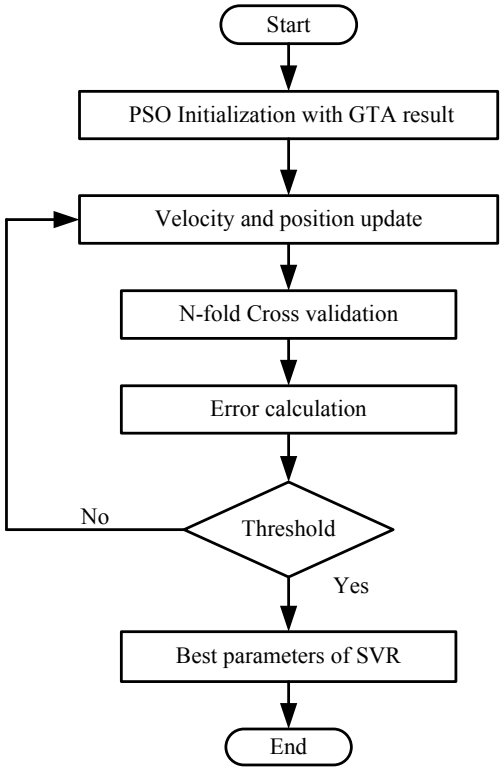


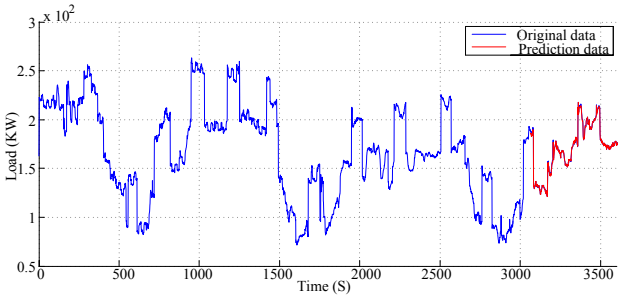
Fig. 3. Flowchart of the designed PSO for the best parameters.

As shown in Fig. 3, the proposed PSO is combined with a cross-validation, and its objective function is similar to (6). For initialization, a number of particles are randomly initialized with given velocities and positions in the solution space. At each iteration, according to the result of the objective function, the velocities of the particles are updated through the historical best position of themselves and the neighbours. Eventually, the best parameters of the SVR can be determined by the particles swarm to the position of the optimal solution. The detail description of the PSO algorithm is shown in Appendix E.

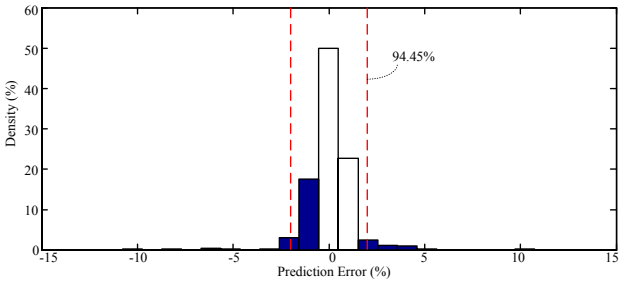
2) Optimal Parameter Determination with PSO: A Distribution System Load Data Prediction: Based on the results in Section IV-A3, the initialization area of the PSO is the blue area in Fig. 2. In this example, $\varepsilon = 0.001$, the PSO is used to determine the best value of C and γ during their lower and upper limits as in Section IV-A3. Considering that the PSO is based on the result of the GTA, the range of velocity is set to be relatively small, as $(0.8, 1.2)$. According to the PSO, the best parameters are determined as $C = 532$ and $\gamma = 977$. With the best parameters, the prediction results are demonstrated as in Fig. 4.

As shown in Fig. 4(a), the prediction result is shown in red curves with triangles, and the original data is shown using blue curves. The training data section is from 0 s to 3,060 s, and the prediction data section is from 3,061 s to 3,600 s. It is noticed that the prediction data curve and original data curve are almost consistent except for several abrupt deviation points. This demonstrates the effectiveness of the proposed prediction approach.

In Fig. 4(b), the mean absolute percentage error (MAPE) is 1.57%, and more than 94% of the errors are accumulated in range $(-2\%, 2\%)$, which also indicates the effectiveness and accuracy of the proposed approach. And some forecast points with relatively large errors demonstrate that the stochastic and abrupt deviating load points in a distribution system are difficult to track and forecast. This also demonstrates that compared to a transmission system, the stochastic characteristic of load deviation is more complicated and widespread in a distribution system.



(a)



(b)

Fig. 4. Prediction result for a distribution system load: (a) original data and prediction data, (b) the percentage error of prediction result

V. Numerical Simulation and Results

The tested data set composes 80 days of load captured from a partner utility's distribution feeder. It includes data from winter (Dec.-Feb.), spring (Mar.-May.), summer (Jun.-Aug.), and autumn (Sep.-Nov.) for 20-days each season. With the sampling rate of 1 Hz, the total data length is 6,912,000. Different from the load data profile in transmission systems as discussed in Section I, there are a lot of abruptly stochastic deviations in the load profile of the distribution system, as shown in Fig. 5. The resolution of the proposed forecasting approach is 1 second. The simulations are executed using a server with 3.60 GHz Intel Xeon CPU and 32 GB RAM. The SVR and PSO are implemented using the LibSVM, MATLAB global optimization toolbox, and parallel computing toolbox.

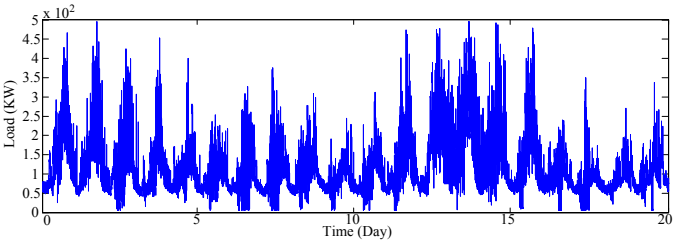


Fig. 5. An example of the collected four season data: 20-day summer load data of a distribution system.

A. Numerical Results

1) Numerical Results on Different Timescales: Considering different control strategies for transient and short-term disturbance, the proposed approach is implemented in different timescales: minutes-ahead and hours-ahead. In our study, the training data is 10 times than the testing data. To evaluate the proposed method comprehensively, the sliding window test is employed to traverse the whole load data. In detail, for example, the sliding window test employed with 5 minutes ahead forecasting can be illustrated as follows.

1) First, the load data section from 1 to 50 minutes is taken as training data to determine the best parameters and build the forecasting model.
2) In the second step, the load data section from 51 to 55 minutes is taken as the test data to evaluate the performance of the forecasting model.
3) In the third step, for the next round forecasting, the training data section moves forward from 6 to 55 minutes, and the test data section moves from 56 to 60 minutes.
4) When the test data section moves to the end of the load data, the sliding window test with 5-minutes forecasting is completed.

Minutes-Ahead Forecasting

In the minutes-ahead forecasting, different forecasting time scales are considered, such as 5-minute-ahead, 10-minute-ahead, 20-minute-ahead, and 40-minute-ahead forecasting. According the forecasting performance demonstrated in Table I, 5-minute-ahead has the best performance and the 40-minute-ahead forecasting has the largest forecasting errors. However, the MAPE of the four types of forecasting are below 2%, and the average MAPE is 1.488%, which indicates that the proposed approach has accurate and robust forecasting performance in minutes-ahead forecasting.

TABLE I
PERFORMANCE OF MINUTES-AHEAD FORECASTING (MAPE(%))

	winter	spring	summer	autumn	Avg.
5-minute-ahead	1.304	1.336	1.312	1.298	1.313
10-minute-ahead	1.453	1.434	1.455	1.477	1.455
20-minute-ahead	1.545	1.578	1.531	1.523	1.544
40-minute-ahead	1.645	1.631	1.637	1.661	1.643

As shown in Fig. 6, it is noticed that more than 75% of the errors are accumulated between $(-2\%, 2\%)$.

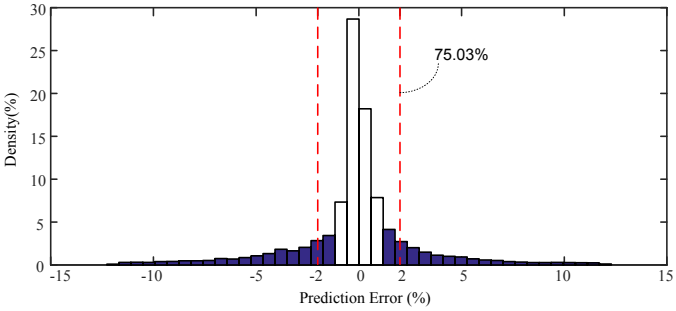


Fig. 6. Percentage error of minutes-ahead scales forecasting.

Hours-Ahead Forecasting

As shown in previous section, the proposed method has very good performance at minutely level. But for some applications, short-term load forecasting also includes hourly time range. For the hours-ahead forecasting, it is considered that there are several time period forecasting scales, such as 1-hour-ahead, 2-hour-ahead, 4-hour-ahead, 8-hour-ahead, and 16-hour-ahead forecasting. According the forecasting performance in Table II, the 1-hour-ahead has the best performance and the 16-hour-ahead forecasting has the largest forecasting errors. However, the MAPE of the five types forecasting are below 3.5%, and the average MAPE is 2.271%. Compared to the forecasting results in the minutes-ahead forecasting, the errors of the hours-ahead forecasting sightly increase. On the other hand, this also illustrates that the hours-ahead forecasting is more complicated than the minutes-ahead forecasting [16], [18].

As shown in Fig. 7, it is noticed that more than 75% of the errors are accumulated between $(-3.9\%, 3.9\%)$.

TABLE II
PERFORMANCE OF HOURS-AHEAD FORECASTING (MAPE(%)).

	winter	spring	summer	autumn	Avg.
1-hour-ahead	1.639	1.647	1.643	1.672	1.650
2-hour-ahead	1.995	1.934	1.981	2.019	1.982
4-hour-ahead	2.169	2.173	2.176	2.123	2.161
8-hour-ahead	2.388	2.372	2.331	2.401	2.373
16-hour-ahead	3.198	3.155	3.172	3.221	3.187

In this paper, the skewness is used to measure the asymmetry of the probability distribution of the forecast errors, which can be defined as following [37]–[39]

$$S_{skw} = \frac{E(\epsilon - \mu)^3}{\sigma^3} \tag{7}$$

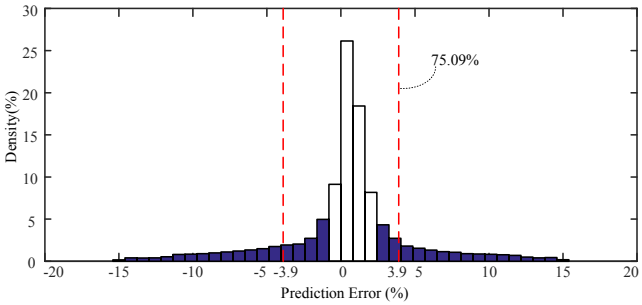


Fig. 7. Percentage error of hours-ahead scales forecasting

where S_{skw} is the skewness, ϵ is the forecast error, μ is the mean of the forecast error, and σ s the standard deviation of the forecast error.

Furthermore, the kurtosis is used to measure the outlier-prone of the probability distribution of the forecast errors, which can be defined as following [37]–[39]

$$S_{krt} = \frac{E(\epsilon - \mu)^4}{\sigma^4} \tag{8}$$

where S_{krt} is the kurtosis, ϵ, μ, and σ are the same as above.

As shown in Table III, the skewness of the minutes-ahead forecasting and hours-ahead are 0.072 and 0.081, which indicates that the proposed approach has a very small asymmetry of the probability distribution of the forecast errors. A few large forecasting errors, which achieve 11.57% result from the abrupt load deviation in the distribution system. The kurtosis of the minutes-ahead forecasting and hours-ahead forecasting are 6.912 and 5.724, which indicate the high accuracy of the proposed approach.

TABLE III
THE SKEWNESS AND KURTOSIS OF THE PROPOSED APPROACH

	skewness	kurtosis
minutes-ahead forecasting	0.072	6.912
hours-ahead forecasting	0.081	5.724

2) Peak Loads Forecasting: As shown in Table IV, the peak load forecasting is investigated with the 1-hour-ahead and 2-hour-ahead forecasting for four seasons data load. Based on [40]–[42], the procedures of peak load forecasting is shown as following:

1) First, the peak load values $\{L^{se}_{p_1}, L^{se}_{p_2}, \cdots, L^{se}_{p_{20}}\}$ and their time indices $\{T^{se}_{p_1}, T^{se}_{p_2}, \cdots, T^{se}_{p_{20}}\}$ are collected, where $se = \{1, 2, 3, 4\}$, which indicates winter, spring, summer and winter. The time range of peak load occurrence is between 11:00 AM to 3:00 PM.
2) In the second step, the historical data are collected to train the SVR models. Specifically, for day 1, the historical data are collected from 0:01 AM to 11:00 AM.
3) In the third step, compare the forecasted loads at time $\{T^{se}_{p_1}, T^{se}_{p_2}, \cdots, T^{se}_{p_{20}}\}$ with the collected peak loads, and compute the forecast errors.

TABLE IV
PERFORMANCE OF PEAK LOADS FORECASTING

	Max Error (%)	Avg. MAPE (%)
1-hour-ahead	6.77	1.615
2-hour-ahead	9.24	1.912

The average MAPE values of the two methods are below 2.0%, which demonstrate the effectiveness of the proposed approach.

B. Compared with Other Methods

In [15]–[18], ARIMA, GA-based SVM, and ANN-based short-term load forecast approaches are presented. In Table I and II, the performances of 20-minute-ahead forecasting and 4-hours-ahead forecasting are in the central positions of minutes- and hour-ahead forecasting, respectively. We use these two evaluations to compare the performance of proposed method with other methods proposed in other literatures. The the forecasting error and time calculation speed of different methods are illustrated in Table V and VI.

Compared to ARIMA, the forecasting approach of ANN- and GA-based SVM are more accurate, but both of them are very time consuming, which take 683.62 s and 1412.7 s in 4-hour-ahead forecasting, respectively. In Table V, the proposed approach has the best performance in both maximum forecasting errors and MAPE, which proves the accuracy of the proposed approach.

In Table VI, the proposed approach takes 12.89 s and 83.53 s in 20-minute-ahead and 4-hour-ahead forecasting, which are much less than the performance of ANN- and GA- based SVM forecasting. Although the ARIMA takes 11.25 s and 77.21 s, it's accuracy are much worse than the proposed approach. Table V and Table VI prove that the proposed approach is a computation-saving and accurate forecasting method.

TABLE V
PERFORMANCE OF DIFFERENT METHODS

Methods	Max. Error (%)	MAPE (%)
ARIMA	31.25	11.21
GA based SVM	21.16	5.27
ANN	25.97	6.62
Proposed	14.11	2.53

TABLE VI
TIME CONSUMPTION OF DIFFERENT METHODS

Methods	20 minutes (S)	4 hours (S)
ARIMA	11.25	77.21
GA based SVM	45.16	1412.7
ANN	40.9	683.62
Proposed	12.89	83.53

VI. CONCLUSION

With the development of modern distribution systems, short-term load forecasting becomes a significant and indispensable issue.

The main contribution of this paper can be concluded as: in this paper, an effective short-term load forecasting approach with high resolution is proposed for the aggregated loads of a small section of distribution feeder load. This type of load forecast will contribute greatly for distributed renewable integration as well as real-time operation of demand side response. Such load profile contains more stochastically abrupt deviations because the behavior of the end users has much bigger impact on distribution systems than it does on transmission systems. Although many SVR based forecasters exist, the proposed two-step hybrid global optimization method can determine the best parameters effectively with acceptable time complexity and computation loads. Other methods, such as ARIMA, GA-based SVM, and ANN-based forecasting methods, are compared to the proposed approach. Considering time consumption and forecasting performance, the proposed approach has the best performance among these methods.

In real-world applications, the load profile has a relationship to other factors, such as temperature, work time, etc.. These factors can provide more information related to load deviation, and the proposed approach is designed to handle multiple parameters and big data non-linear regression. If these factors considered, the proposed approach can provide better forecasting results. Considering the parallel computation model, Mapreduce, which is implemented in many different computation frames such as Apache Hadoop, Apache Spark, the time consumption can be further reduced. In the future, it is expected that the proposed approach can be adapted to cooperate with existing wind and solar forecasting methods and facilitate better integration of renewable energy resources into distribution systems.

APPENDIX A
DETAIL PROBLEM FORMULATION

The flowchart of the proposed approach is shown in Fig. 1. At the beginning, the collected historical load data vector $\mathbf{L} = [L_1 \ \cdots L_i \ \cdots L_n]$ is processed by the data preprocessing algorithm (Block 1), where L_i denotes the loads in a distribution system at time i, and $i \in \{1 \ \cdots \ n\}$. In Block 1, the original data set $\mathbf{L}$ will be preprocessed and mapped into a vector $\mathbf{L}' = [L'_1 \ \cdots \ L'_i \ \cdots \ L'_n]$ to reduce the excessive deviation, where L'_i is the mapped load at time i. $\hat{\mathbf{L}}$ is the forecasting load data vector, where $\hat{L}_{i_1}$ is the i_1th element of the vector $\hat{\mathbf{L}}$, where $i_1 \in \{n+1, \ \cdots, \ n+m\}$.

The parameter optimization of the SVR contains a two-step optimization approach. It is defined that is a parameter vector of the SVR, and $_j$ is the jth parameter, where $j \in \{1, 2, \cdots, p\}$. $\bar{}$ and $\_$ are the upper limit vector and lower limit vector of , $\bar{}_j$ and $\_{}_j$ are the jth elements of $\bar{}$ and $\_$, respectively. $\mathbf{\Delta}$ indicates a searching step vector of the GTA, and Δ_j is the jth element of $\mathbf{\Delta}$.

At the beginning of the parameter optimization, all the parameters are initiated (Block 2). In Block 3, based on the searching step vector $\mathbf{\Delta}$, the parameter vector , the upper limit vector $\bar{}$, and lower limit vector $\_$, a multi-dimensional vector $\mathbf{\Lambda}$ is computed to build a traverse vector $\mathbf{H}$, which contains all the combinations of different parameters

in $\mathbf{\Lambda}$. Combined with the cross-validation in Block 4, the GTA traverses down their performance to narrow the possible solution from the global solution space R^p into a local solution space R_1^p, where $R_1^p \subset R^p$. In the GTA, all the elements in $\mathbf{\Delta}$ are set relatively large to reduce the time consumption yet not neglect the local possible solution space R_1^p.

In the second step of the parameter optimization, the particles of the PSO are initialized with the result of the GTA in the solution space (Block 6). Combined with cross-validation in Block 7, at each iteration, the position vector $\boldsymbol{\alpha}_{i_4}$, velocity vector $\boldsymbol{\nu}_{i_4}$, and the best position vector $\boldsymbol{\eta}_{i_4}$ are updated with the result of the objective function, where $i_4 \in \{1, 2, 3, \cdots, n_{PSO}\}$, n_{PSO} is the number of particles in the PSO, and is a symbol to indicate that the particles of the PSO are initialized in the GTA solution space of the SVR parameters. In the PSO, the elements of the velocity vector $\boldsymbol{\alpha}_{i_4}$ are set relatively small to determine the best parameters in the local solution space R_1^p accurately. The best parameter vector H_b can be determined in Block 9. In the next step, $\hat{\mathbf{L}}$ is forecasted by the SVR forecaster with the best parameter vector H_b in Block 10. In Block 11, the report of the numerical result is generated and compared to other approaches.

In real-world applications, because the traverse result corresponding to each element in $\mathbf{H}$ can be calculated independently, a parallel computation of GTA is used to reduce consumption time. If there are several possible local solution spaces, $R_1^p, R_2^p, \cdots, R_{nr}^p$, determined by the GTA, a parallel computation of the PSO can be used to determine several best parameter vectors $H_b^{p_1}, H_b^{p_2}, \cdots, H_b^{p_{nr}}$ for each small possible solution space, respectively, where $R_{j_1}^p \in R^p$, $j_1 \in \{1, 2, \cdots, nr\}$. Finally, comparing all the parameter vectors, a best parameter vector H_b can be determined for the SVR.

After the SVR-based forecaster is built, the future load data $\hat{\mathbf{L}}$ can be forecast from historical data $\mathbf{L}$.

Appendix B
Data Preprocessing

To eliminate the excessively deviating instances, a mapping algorithm is used to map the load profile $\mathbf{L}$ to a small range $\mathbf{L}'$. The algorithm is given by

$$K_{pre} = (L'_{\max} - L'_{\min})/(L_{\max} - L_{\min}) \tag{9}$$

$$L'_i = K_{pre}(L_i - L_{\min}) + L'_{\min} \tag{10}$$

where K_{pre} is a mapping ratio from the original load $\mathbf{L}$ to the small range $\mathbf{L}'$, $L_{\max}$ and $L_{\min}$ indicate the maximum value and minimum value of $\mathbf{L}$, respectively. $L'_{\max}$ and $L'_{\min}$ indicate the maximum value and minimum value of $\mathbf{L}'$, respectively. In this paper, $\mathbf{L}'$ is normalized to range $(0, 1)$, which means $L'_{\max}$ and $L'_{\min}$ are set to 1 and 0, respectively. With the inverse function of the mapping algorithm (1) and (2), the forecasting result $\hat{\mathbf{L}}$ can be computed.

Appendix C
Procedure for Building the Traversing Vector

The traversing vector $\mathbf{H}$ is built as follows:

1) For every different vector Λ_j, a parameter β_{j,m''_j} is chosen, where j and m''_j indicate that the parameter β_{j,m''_j} is the m''_jth element chosen from vector Λ_j, $j \in \{1, 2, \cdots, p\}$ and $m''_j \in \{1, 2, \cdots, m_j\}$.
2) H_{j_2} is the j_2th element of $\mathbf{H}$, which can be built as a combination of different parameters $[\beta_{1,m''_1} \cdots \beta_{j,m''_j} \cdots \beta_{p,m''_p}]$, where $j_2 \in \{1, 2, \cdots, m_1 \times m_2 \times, \cdots, \times m_p\}$.
3) After every element vector of $\mathbf{H}$ is computed, the traversing vector $\mathbf{H}$ is finished.

In addition, because the three parameters γ, C, and ε are chosen, the dimension of vector $\mathbf{\Lambda}$ is reduced, and $\mathbf{\Lambda} \in R^{m_1 \times m_2 \times m_3}$. Therefore, the total number of the element vector H_{j_2} in $\mathbf{H}$ is reduced, $j_2 \in \{1, 2, \cdots, m_1 \times m_2 \times m_3\}$, which means the computation complexity of GTA is reduced.

For example, $m_1 = 2$, $m_2 = 2$, $m_3 = 1$, then, the number of element in $\mathbf{H}$ is $2 \times 2 \times 1 = 4$. $H_1 = [\gamma_1\ C_1\ \varepsilon_1]$, $H_2 = [\gamma_1\ C_2\ \varepsilon_1]$, $H_3 = [\gamma_2\ C_1\ \varepsilon_1]$, and $H_4 = [\gamma_2\ C_2\ \varepsilon_1]$.

Appendix D
The GTA Procedure

The GTA is shown in Algorithm 1, which aims to traverse the global solution space into one or several local spaces for the PSO in the next step. In the first step, the three parameters are initialized and the traversing vector $\mathbf{H}$ is built. In the second step, for each element vector H_{j_2} in $\mathbf{H}$, the computation of the corresponding R_{CV} is independent and can be computed in parallel, which can further reduce the computation time. In the last step, if multiple local solution spaces are determined with the contour map, all of them are selected for the PSO step in parallel.

Algorithm 1 GTA for Parameter Optimization

Objective: Transfer the global optimization problem to one or several local optimization problems.

Initialization: Initialize γ, C, and ε; then compute Λ_j with formula (5), and build the traverse vector $\mathbf{H}$.

Grid Traverse Searching: For the element factor H_{j2}, $H_{j2} \in \mathbf{H}$, $j_2 \in \{1, 2, \cdots, m_1 \times m_2 \times m_3\}$, the R_{CV} is computed with formula (6).

Determine Local Solution Space: With the generated contour map, the local solution space with minimum R_{CV} is selected for next step of optimization.

Appendix E
The PSO Algorithm

Based on the result of the GTA, in the initialization part of the PSO, the i_4th particle can be designed as follows,

$$\boldsymbol{\alpha}_{i_4} = [\alpha_{i_4,1}\alpha_{i_4,2} \cdots \alpha_{i_4,n_{OBJPSO}}] \tag{11}$$

$$\boldsymbol{\nu}_{i_4} = [\nu_{i_4,1}\nu_{i_4,2} \cdots \nu_{i_4,n_{OBJPSO}}] \tag{12}$$

$$\boldsymbol{\eta}_{i_4} = [\eta_{i_4,1}\eta_{i_4,2} \cdots \eta_{i_4,n_{OBJPSO}}] \tag{13}$$

where $\boldsymbol{\alpha}_{i_4}$, $\boldsymbol{\nu}_{i_4}$, and $\boldsymbol{\eta}_{i_4}$ indicate the vector of the position, velocity, and best position of the i_4th particle, receptively.

According to the analysis above, = $[\gamma\ C\ \varepsilon]$; and n_{OBJPSO} is the dimension of the solution space, as a result, n_{OBJPSO} = 3 in this paper. $i_4 = 1, 2, 3, \cdots, N_{PSO}$, N_{PSO} is the number of the particles in the PSO.

At each iteration, the new velocity of the i_4th particle is determined by its own velocity, the best historical position and the global best position.

$$\begin{aligned}\boldsymbol{\nu}_{i_4}(t) =& \boldsymbol{\nu}_{i_4}(t-1) + \varphi_1\theta_1(\boldsymbol{\eta}_{i_4} - \boldsymbol{\alpha}_{i_4}(t-1)) \\ &+ \varphi_2\theta_2(\boldsymbol{\eta}_g - \boldsymbol{\alpha}_{i_4}(t-1))\end{aligned} \tag{14}$$

where φ_1 and φ_2 are the two positive acceleration coefficients, θ_1 and θ_2 are two independently uniformly distributed random variables with range $(0, 1)$, $\boldsymbol{\eta}_{i_4}$ is the vector of the best historical position of the i_4th particle, and $\boldsymbol{\eta}_g$ is the best position vector among all particles.

At each iteration, the position of the i_4th particle is determined as follows:

$$\boldsymbol{\alpha}_{i_4}(t) = \boldsymbol{\alpha}_{i_4}(t-1) + \boldsymbol{\nu}_{i_4}(t) \tag{15}$$

where $\boldsymbol{\alpha}_{i_4}(t)$ is the position vector of the i_4th particle at time t, $\boldsymbol{\nu}_{i_4}(t)$ is the velocity vector of the i_4th particle at time t, and = $[\gamma\ C\ \varepsilon]$.

After the positions and velocities of all the particles are updated at each iteration, the performance of the parameters are computed with the designed cross-validation. A cross-validation is used with the PSO similarly as that shown in Section IV-A2, and the objective function is similar to (6).

REFERENCES

[1] J. M. Carrasco, L. G. Franquelo, J. T. Bialasiewicz, E. Galván, R. P. Guisado, M. A. Prats, J. I. León, and N. Moreno-Alfonso, "Power-electronic systems for the grid integration of renewable energy sources: A survey," *IEEE Transactions on Industrial Electronics*, vol. 53, no. 4, pp. 1002–1016, 2006.

[2] Y. Gu, H. Jiang, Y. Zhang, and D. W. Gao, "Statistical scheduling of economic dispatch and energy reserves of hybrid power systems with high renewable energy penetration," in *2014 48th Asilomar Conference on Signals, Systems and Computers*, 2014, pp. 530–534.

[3] H. Jiang, J. J. Zhang, W. Gao, and Z. Wu, "Fault detection, identification, and location in smart grid based on data-driven computational methods," *IEEE Transactions on Smart Grid*, vol. 5, pp. 2947 – 2956, 2014.

[4] F. He, Y. Gu, J. Hao, J. J. Zhang, J. Wei, and Y. Zhang, "Joint real-time energy and demand-response management using a hybrid coalitional-noncooperative game," in *2015 49th Asilomar Conference on Signals, Systems and Computers*. IEEE, 2015, pp. 895–899.

[5] F. Ding, P. Li, B. Huang, F. Gao, C. Ding, and C. Wang, "Modeling and simulation of grid-connected hybrid photovoltaic/battery distributed generation system," in *CICED 2010 Proceedings*. IEEE, 2010, pp. 1–10.

[6] X. Fang, S. Misra, G. Xue, and D. Yang, "Smart grid - the new and improved power grid: a survey," *IEEE Communications Surveys and Tutorials*, vol. 14, pp. 944–980, 2012.

[7] H. Jiang, J. J. Zhang, D. W. Gao, Y. Zhang, and E. Muljadi, "Synchrophasor based auxiliary controller to enhance power system transient voltage stability in a high penetration renewable energy scenario," in *2014 IEEE Symposium Power Electronics and Machines for Wind and Water Applications (PEMWA)*, 2014, pp. 1–7.

[8] E. Dall'Anese, H. Zhu, and G. B. Giannakis, "Distributed optimal power flow for smart microgrids," *IEEE Transactions on Smart Grid*, vol. 4, no. 3, pp. 1464–1475, 2013.

[9] T. Hong, P. Pinson, and S. Fan, "Global energy forecasting competition 2012," *International Journal of Forecasting*, vol. 30, no. 2, pp. 357–363, 2014.

[10] J. Xiao, X.-d. Guo, and L.-q. Bai, "An improved model of total supply capability for distribution systems," in *Power and Energy Engineering Conference (APPEEC), 2012 Asia-Pacific*. IEEE, 2012, pp. 1–4.

[11] Q. Wang, J. D. McCalley, T. Zheng, and E. Litvinov, "A computational strategy to solve preventive risk-based security-constrained opf," *IEEE Transactions on Power Systems*, vol. 28, no. 2, pp. 1666–1675, 2013.

[12] H. Jiang, L. Huang, Y. Zhang, J. J. Zhang, and D. W. Gao, "Spatial-temporal characterization of synchrophasor measurement systems - A big data approach for smart grid system situational awareness," in *2014 Conference Record of the Forty Eighth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, 2014, pp. 750–754.

[13] P.-F. Pai and W.-C. Hong, "Support vector machines with simulated annealing algorithms in electricity load forecasting," *Energy Conversion and Management*, vol. 46, no. 17, pp. 2669–2688, 2005.

[14] D. Niu, Y. Wang, and D. D. Wu, "Power load forecasting using support vector machine and ant colony optimization," *Expert Systems with Applications*, vol. 37, no. 3, pp. 2531–2539, 2010.

[15] P.-F. Pai and W.-C. Hong, "Forecasting regional electricity load based on recurrent support vector machines with genetic algorithms," *Electric Power Systems Research*, vol. 74, no. 3, pp. 417–425, 2005.

[16] C. Lu, H.-T. Wu, and S. Vemuri, "Neural network based short term load forecasting: A reivew and evaluation," *IEEE Transactions on Power Systems*, vol. 8, no. 1, pp. 336–342, 1993.

[17] A. Khotanzad, R.-C. Hwang, A. Abaye, and D. Maratukulam, "An adaptive modular artificial neural network hourly load forecaster and its implementation at electric utilities," *IEEE Transactions on Power Systems*, vol. 10, no. 3, pp. 1716–1722, 1995.

[18] S.-J. Huang and K.-R. Shih, "Short-term load forecasting via ARMA model identification including non-gaussian process considerations," *IEEE Transactions on Power Systems*, vol. 18, no. 2, pp. 673–679, 2003.

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

[20] J. W. Taylor, "Short-term electricity demand forecasting using double seasonal exponential smoothing," *Journal of the Operational Research Society*, vol. 54, no. 8, pp. 799–805, 2003.

[21] J. Taylor, A. Maitra, M. Alexander, D. Brooks, and M. Duvall, "Evaluations of plug-in electric vehicle distribution system impacts," in *2010 IEEE Power and Energy Society General Meeting*, 2010, pp. 1–6.

[22] K. Clement-Nyns, E. Haesen, and J. Driesen, "The impact of charging plug-in hybrid electric vehicles on a residential distribution grid," *IEEE Transactions on Power Systems*, vol. 25, no. 1, pp. 371–380, 2010.

[23] H. Jiang, X. Dai, W. Gao, J. Zhang, Y. Zhang, and E. Muljadi, "Spatial-temporal synchrophasor data characterization and analytics in smart grid fault detection, identification and impact causal analysis," *IEEE Transactions on Smart Grid*, vol. 7, no. 5, pp. 2525–2536, 2016.

[24] M. Cui, D. Ke, Y. Sun, D. Gan, J. Zhang, and B.-M. Hodge, "Wind power ramp event forecasting using a stochastic scenario generation method," *IEEE Transactions on Sustainable Energy*, vol. 6, no. 2, pp. 422–433, 2015.

[25] Z. Tang and K. K. Bagchi, "Globally convergent particle swarm optimization via branch-and-bound," *Computer and Information Science*, vol. 3, no. 4, pp. 60–71, 2010.

[26] F. Van Den Bergh, "An analysis of particle swarm optimizers," Ph.D. dissertation, University of Pretoria, 2006.

[27] S. Kotsiantis, D. Kanellopoulos, and P. Pintelas, "Data preprocessing for supervised leaning," *International Journal of Computer Science*, vol. 1, no. 2, pp. 111–117, 2006.

[28] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and Computing*, vol. 14, no. 3, pp. 199–222, 2004.

[29] C. J. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 121–167, 1998.

[30] C.-C. Chang and C.-J. Lin, "Libsvm: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 3, p. 27, 2011.

[31] A. Shashua, "Introduction to machine learning: Class notes 67577," *arXiv preprint arXiv:0904.3664*, 2009.

[32] M. Jiang, S. Jiang, L. Zhu, Y. Wang, W. Huang, and H. Zhang, "Study on parameter optimization for support vector regression in solving the inverse ECG problem," *Computational and mathematical methods in medicine*, vol. 2013, 2013.

[33] A. R. Webb, *Statistical pattern recognition*. John Wiley & Sons, 2003.

[34] S. Makridakis, S. C. Wheelwright, and R. J. Hyndman, *Forecasting methods and applications*. John Wiley & Sons, 2008.

[35] S. Fan and L. Chen, "Short-term load forecasting based on an adaptive hybrid method," *IEEE Transactions on Power Systems*, vol. 21, no. 1, pp. 392–401, 2006.

[36] Y. Del Valle, G. K. Venayagamoorthy, S. Mohagheghi, J.-C. Hernandez, and R. G. Harley, "Particle swarm optimization: basic concepts, variants and applications in power systems," *IEEE Transactions on Evolutionary Computation*, vol. 12, no. 2, pp. 171–195, 2008.

[37] K. V. Mardia, "Measures of multivariate skewness and kurtosis with applications," *Biometrika*, vol. 57, no. 3, pp. 519–530, 1970.

[38] ——, "Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies," *Sankhy$\bar{a}$: The Indian Journal of Statistics, Series B*, pp. 115–128, 1974.

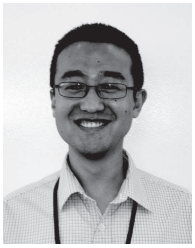
[39] E. Jondeau and M. Rockinger, "Conditional volatility, skewness, and kurtosis: existence, persistence, and comovements," *Journal of Economic Dynamics and Control*, vol. 27, no. 10, pp. 1699–1737, 2003.

[40] Y.-Y. Hsu and C.-C. Yang, "Design of artificial neural networks for short-term load forecasting. part 2: Multilayer feedforward networks for peak load and valley load forecasting," in *IEE Proceedings-Generation, Transmission and Distribution*, vol. 138, no. 5. IET, 1991, pp. 414–418.

[41] L. Saini and M. Soni, "Artificial neural network based peak load forecasting using levenberg-marquardt and quasi-newton methods," in *IEE Proceedings-Generation, Transmission and Distribution*, vol. 149, no. 5. IET, 2002, pp. 578–584.

[42] N. Amjady, "Short-term hourly load forecasting using time-series modeling with peak load estimation capability," *IEEE Transactions on Power Systems*, vol. 16, no. 3, pp. 498–505, 2001.

Huaiguang Jiang received his B.E. degree in electrical engineering from the National University of Defence Technology, Changsha, China, in 2007, his M.S. degree from the electrical engineering in University of Electronic Science and Technology of China, Chengdu, China, in 2010, and his Ph.D in Electrical Engineering, University of Denver, Denver, Colorado, USA, in 2015. Currently, his is with the National Renewable Energy Laboratory in Golden, Colorado. His research interests are phasor measurement unit applications, renewable energy integration, smart grid, signal processing, time-frequency analysis, big data, and machine learning.



Yingchen ("Y.C.") Zhang (M07) received his B.S.E.E. from Tianjin University, Tianjin, China, in 2003 and his Ph.D. in electrical engineering from Virginia Polytechnic Institute and State University, Blacksburg, in 2010. He is currently with the National Renewable Energy Laboratory in Golden, Colorado, USA. His research interests include power system stability with the large-scale integration of renewable energies, power system wide-area monitoring, and phasor measurement unit applications for renewable integrations.

Eduard Muljadi (M82-SM94-F10) received his Ph. D. (in electrical engineering) from the University of Wisconsin, Madison. From 1988 to 1992, he taught at California State University, Fresno, California. In June 1992, he joined the National Renewable Energy Laboratory in Golden, Colorado. His current research interests are in the fields of electric machines, power electronics, and power systems in general with emphasis on renewable energy applications. He is member of Eta Kappa Nu, Sigma Xi, and a Fellow of the IEEE. He is involved in the activities of the IEEE Industry Application Society (IAS), Power Electronics Society, and Power and Energy Society (PES). He is currently a member of various committees of the IAS, a member of the Working Group on Renewable Technologies and Dynamic Performance Wind Generation Task Force of the PES, and an editor of the *IEEE Transactions on Energy Conversion*. He holds two patents in power conversion for renewable energy.

Jun Jason Zhang (M09) received his B.E. and M.E. degrees in electrical engineering from Huazhong University of Science and Technology, Wuhan, China, in 2003 and 2005, respectively, and his Ph.D. in electrical engineering from Arizona State University, Tempe, in 2008. He is currently an assistant professor of electrical and computer engineering at the University of Denver. His research interests are in the areas of agile sensing, signal processing and implementation, time-varying system modeling, and their applications in intelligent power and energy systems.

David Wenzhong Gao (M02SM03) received his M.S. and Ph.D. degrees in electrical and computer engineering, specializing in electric power engineering, from Georgia Institute of Technology, Atlanta, USA, in 1999 and 2002, respectively. He is now an associate professor in the Department of Electrical and Computer Engineering, University of Denver, Colorado, USA. His current teaching and research interests include renewable energy and distributed generation, microgrid, smart grid, power system protection, power electronics applications in power systems, power system modeling and simulation, and hybrid electric propulsion systems. He is an editor of *IEEE Transactions on Sustainable Energy* and an associate editor of *IEEE Journal of Emerging and Selected Topics in Power Electronics*. He is the General Chair for The IEEE Symposium on Power Electronics and Machines in Wind Applications (PEMWA 2012) and North American Power Symposium 2016.