Exploring Key Weather Factors From Analytical Modeling Toward Improved Solar Power Forecasting

Jianxiao Wang, *Student Member, IEEE*, Haiwang Zhong, *Member, IEEE*, Xiaowen Lai, Qing Xia, *Senior Member, IEEE*, Yang Wang, and Chongqing Kang, *Fellow, IEEE*

***Abstract*—Accurate solar power forecasting plays a critical role in ensuring the reliable and economic operation of power grids. Most of existing literature directly uses available weather conditions as input features, which might ignore some key weather factors and the coupling among weather conditions. Therefore, a novel solar power forecasting approach is proposed in this paper by exploring key weather factors from photovoltaic (PV) analytical modeling. The proposed approach is composed of three engines: 1) analytical modeling of PV systems; 2) machine learning methods for mapping weather features with solar power; and 3) a deviation analysis for solar power forecast adjustment. In contrast to the existing research that directly uses available weather conditions, this paper explores the physical knowledge from PV models. Different irradiance components and PV cell temperatures are derived from PV analytical modeling. These weather features are used to reformulate the input of machine learning methods, which helps achieve a better forecasting performance. Moreover, based on the historical forecasting deviations, a compensation term is presented to adjust the solar power forecast. Case studies based on measured datasets from PV systems in Australia demonstrate that the forecasting performance can be highly improved by taking advantage of the key weather features derived from PV models.**

***Index Terms*—Analytical modeling, solar power forecasting, deviation analysis, weather knowledge.**

NOMENCLATURE

t	Time index
T	Number of historical observations
L	Number of extracted principal components
H	Number of forecasting periods
P_0	Reference power on standard weather conditions
E_0	Reference irradiance (1000 W/m^2)
T_0	Reference temperature (25°C)
c_{SF}	Soiling factor of the PV system
c_T	Temperature factor of the PV system
P_{mp}	Solar power at the maximum power point (MPP)
E_{POA}	Plane of array irradiance
T_C	PV cell temperature
E_b	Beam component of E_{POA}
E_g	Ground-reflected component of E_{POA}
E_d	Sky-diffuse component of E_{POA}
E_{DNI}	Direct normal irradiance
E_{GHII}	Global horizontal irradiance
E_{DHI}	Diffuse horizontal irradiance
$\theta_{T,array}$	Tilt of the PV plane
$f_{DNI}(\cdot)$	Estimation function of E_{DNI}
$f_{DHI}(\cdot)$	Estimation function of E_{DHI}
T_A	Ambient temperature
V_W	Wind speed
$\mathbf{X_C}$	Key weather feature matrix
$\mathbf{X_I}$	Input feature matrix
$\mathbf{X_S}$	Standardized input feature matrix
$\mathbf{X_P}$	Historical principal component matrix
$\mathbf{X_{PF}}$	Principal component matrix for forecasting periods
$\mathbf{X}[i,j]$	(i,j)th entry of matrix $\mathbf{X}$
$\mathbf{X}[\cdot,j]$	jth column of matrix $\mathbf{X}$
$\mathbf{X}[i,\cdot]$	ith row of matrix $\mathbf{X}$
$MD(\cdot)$	Manhattan distance
$\widehat{\mathbf{P}}_{\mathbf{mp}}$	Forecasted solar power vector
$\Delta\mathbf{P_{mp}}$	Solar power compensation vector
$\widehat{\mathbf{P}}^*_{\mathbf{mp}}$	Forecasted solar power vector after adjustment
P_C	Capacity of the PV system.

Manuscript received May 20, 2017; revised August 29, 2017; accepted October 12, 2017. Date of publication October 23, 2017; date of current version February 18, 2019. This work was supported in part by the National Key Research and Development Program of China under Grant 2016YFB0900103, in part by the State Grid Corporation of China under the project titled "Risk Quantization and Active Control for Power Grid Operations Considering Large-scale Meteorological Data," and in part by the Junior Fellowships for Advanced Innovation Think-tank Programme, China Association for Science and Technology under Grant DXB-ZKQN-2017-037. Paper no. TSG-00699-2017. *(Corresponding author: Haiwang Zhong.)*

The authors are with the State Key Laboratory of Power Systems, Department of Electrical Engineering, Tsinghua University, Beijing 100084, China (e-mail: wang-jx14@mails.tsinghua.edu.cn; zhonghw@mail.tsinghua.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSG.2017.2766022

I. INTRODUCTION

FACED with severe environmental challenges and energy shortages brought on by the massive consumption of fossil fuels, it has become a global consensus to develop renewable and sustainable energy [1], [2]. Solar energy is one of the most promising and fastest-growing alternative energy resources that has been widely deployed across the world. Up until 2016, the total installed capacity of global photovoltaic (PV) generation has reached 303 GW, increasing by 33.48% compared with that in 2015. China has been managing one of the largest solar industries in the world, with

77.42 GW installed capacity. The four countries with largest PV markets in 2016 are China (34.24 GW), the United States (13 GW), Japan (10.5 GW) and India (4 GW) [3]. Moreover, the PV industry, the PV market and the global installed capacity are expected to maintain sustainable and rapid growth in subsequent years [4].

While the world-wide deployment of solar energy contributes to a more sustainable future, the intermittent and volatile nature of solar power poses significant challenges to the reliable and economic operation of power grids [5]. On one hand, the intermittency of large-scale solar power may cause an imbalance between power supply and demand, influencing the reliability of power grids. On the other hand, more traditional generation resources for rapid ramping and reserve should be scheduled to smooth out the fluctuations in solar power, decreasing the efficiency of generation resources [6]. Thus, it is imperative to improve the forecast accuracy of solar power. In existing literature, solar power forecasting methods can be classified from different perspectives. For example, based on different forecasting engines, the methods can be categorized into machine learning and analytical modeling [7].

In machine learning methods, the relationships between solar power and weather conditions, e.g., surface irradiance, ambient temperature and wind speed, are learned from training with a large amount of historical data. Most of the existing literature focuses on improving the learning ability of the forecasting engine. However, the weather conditions from the numerical weather prediction (NWP) are *directly* used as the input of machine learning, without preprocessing the coupling among these weather conditions.

In [4], a weather-based method for day-ahead hourly forecasting of PV power is presented. That method comprises classification, training and forecasting stages and is validated using real-world historical data of ambient temperature, precipitation and solar irradiance. In [12], a data-driven forecast model is proposed to improve the accuracy of a short-term PV power forecast. The autoregressive method is adopted to forecast 1-h-ahead solar power considering solar irradiance. In [14], a similar-day detection engine is proposed to find several days that are similar to the target forecast day considering irradiance, temperature and wind speed. Then, the forecast accuracy is validated using four different forecasting engines. In [15], support vector machine (SVM) is used to learn the mapping between weather conditions from satellite images and solar power. The weather conditions include wind speed, cloud coverage and irradiance. In [16], a higher-order Markov chain is adopted to forecast the probability distribution of PV power. Ambient temperature and solar irradiance measurements are used to classify different operating conditions of a PV system. In [17], a novel weighted Gaussian process regression approach is proposed, such that data samples with higher outlier have a lower weight. In [18], aerosol index is considered as an additional input parameter of solar power forecast. The case studies show that the linear correlation between aerosol index and solar radiation helps improve forecasting accuracy. In [19], a forecasting framework is proposed to explore information from a grid of numerical weather predictions applied to both wind and solar energy. Compared to a model that only considers one NWP point for a specific location, a grid of weather predictions can greatly reduce the forecasting errors.

To explore the physical relationships between solar power and weather conditions, analytical modeling methods are proposed. In analytical modeling, solar power is theoretically modeled as a function of different weather conditions. Then the parameters in the function are regressed according to field experiments. Most of existing literature focuses on the accurate modeling of a PV system. However, PV modeling is quite complex, due to nonlinear and time-varying parameters in a practical PV system. The analytical modeling-based regression might not have satisfactory performance especially on rainy or cloudy days when weather conditions frequently vary.

In [8], a modified current-voltage relationship for the single-diode model is presented. The model begins with a single solar cell and is expanded to a PV module, then to a PV array. The parameters of the model are regressed with experimental data under different levels of irradiance and temperature. In [9], an empirical model is formulated between solar power and weather conditions, including irradiance, ambient temperature and humidity. Then, the parameters of that model are regressed using actual observations of a PV module. In addition, some recent studies investigate irradiance and temperature modeling of a PV array [10], [11].

Some ensemble methods have been proposed to take advantage of both machine learning and analytical modeling methods. In [20], different weather conditions are respectively forecasted using spatial and temporal time series methods. Then solar power is calculated using the forecasted weather conditions based on an empirical model. However, the ensemble method in [20] still highly relies on the accuracy of PV models. In [21], an ensemble method is proposed, which has three forecasting engines: an autoregressive model, gradient boosting regression model and an empirical PV model. The three engines are separately used to forecast solar power. However, the physical knowledge provided by PV models is not thoroughly used in the forecasting engines.

The literature review is summarized in TABLE I.

According to existing forecasting methods, there remain three issues to be addressed: i) The PV modeling is over-simplified that some *important* weather conditions might be overlooked. For example, solar power is related to different irradiance components, including beam, diffuse and reflection irradiance. ii) The weather measurements from NWP serve as *direct* inputs of machine learning methods, ignoring the correlations between different weather measurements. iii) The nonlinearity between weather conditions and solar power have not been thoroughly explored from PV analytical models. According to PV analytical modeling, some important weather features and the coupling among weather conditions cannot be directly measured or predicted from NWP. Instead, these key weather factors need to be theoretically derived from PV models. Without making full use of these key factors, the ability of machine learning methods cannot be efficiently deployed.

Analytical modeling of PV systems can provide physical knowledge for solar power forecasting, and machine learning

TABLE I
LITERATURE REVIEW ON SOLAR POWER FORECASTING METHODS

Method	Weather feature	Forecasting engine	PV modeling
Machine learning	Ambient temperature[4] [16], cell temperature[14], precipitation[4], solar irradiance[4] [12] [15] [16], wind speed[14], surface irradiance[14], cloud cover[14] [15]	SVM[4] [14] [15], persistence[12][13], K-nearest neighbor (KNN)[14], artificial neural network (ANN) [14], weighted KNN[14], Markov chain[16]	/
Analytical modeling	Ambient temperature[8] [9] [11], cell temperature[8], solar irradiance[8] [9] [11], humidity[9] [10] [11], precipitation[10], cloud cover[10], wind speed[11]	Nonlinear regression[9] [10] [11]	Single-diode model[8] [9], empirical model[10] [11]
Ensemble method	Surface irradiance[20] [21], ambient temperature[20] [21], wind speed[20] [21]	Persistence[20], autoregression[20] [21], ANN[20], compressive spatio-temporal forecasting[20], gradient boosting regression[21]	Empirical model[20] [21]

methods can capture the nonlinear and time-varying parameters of PV systems. To fully deploy the advantages of both methods, a novel solar forecasting approach is proposed in this paper. The proposed approach is comprised of three engines: i) analytical modeling of PV systems; ii) machine learning methods to map weather features and solar power; and iii) a deviation analysis for solar power forecast adjustment. The PV analytical modeling can generate key weather features, and then the machine learning methods take advantage of the key features to improve the forecasting accuracy.

The major contributions of this paper are threefold:

1) The weather knowledge related to solar power is explored from an analytical model. The relationships between solar power and key weather features are deduced.

2) The key weather factors provided by PV analytical modeling are used to reformulate input features, which improves the performance of the machine learning methods.

3) A deviation analysis is conducted to improve the forecasting accuracy. Based on the historical forecasting deviations, the forecasted solar power can be adjusted by a compensation term.

II. THE PROPOSED APPROACH

A. Structure of This Paper

The structure of the proposed solar power forecasting approach is shown in Fig. 1. The basic idea is to reformulate the input of machine learning methods by using the key weather features from analytical modeling. By this means, the physical knowledge from analytical modeling can help improve the performance of the machine learning methods.

In the analytical modeling module, the key weather factors related to solar power are theoretically deduced. Given weather predictions, different components of irradiance and PV cell temperature can be calculated. Then, the key weather features can be derived from the irradiance and the cell temperature. As these key weather features have strong correlations with solar power, the key features and other available weather predictions are combined as the input of machine learning methods. This module will be elaborated in Section II-B.

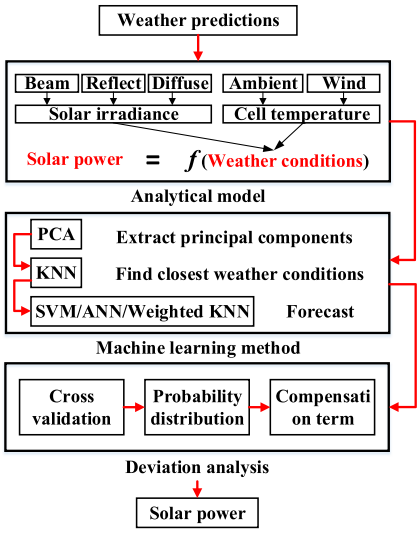

Fig. 1. The structure of the proposed solar power forecasting approach.

In the machine learning module, principal component analysis (PCA) is used for feature extraction. K-nearest neighbor (KNN) is employed to classify a forecasting period into the historical periods with the closest weather conditions. To demonstrate the performance of the analytical modeling integrated with different forecasting engines, three well-known classic forecasting engines [14], i.e., SVM, ANN and weighted KNN, are applied respectively. This module will be elaborated in Section II-C.

In the deviation analysis module, cross-validation is used to obtain the distribution of forecast errors. Then the forecasted solar power can be adjusted by a compensation term. This module will be elaborated in Section II-D.

The major contribution of this paper is to improve forecasting performance by taking advantage of the key weather factors derived from PV analytical modeling. The forecasting engine itself is not the focus of this paper. In the case studies, the solar power forecasting with and without the key weather factors will be compared to demonstrate the impacts brought on by PV analytical modeling.

As the performance of the proposed method highly depends on the accuracy of NWP data, the proposed method can be applied in the hour/day-ahead solar power forecasting.

B. Analytical Modeling of Photovoltaic Systems

The analytical modeling of PV systems provided by the National Renewable Energy Laboratory (NREL) is adopted in this paper [22]. It has been widely used as an online PV performance application. The model is show as follows:

$$P_{mp} = c_{STD} E_{POA} \cdot [1 + c_T (T_C - T_0)], \quad (1)$$

where $c_{STD} = P_0 c_{SF}/E_0$, $c_{SF} \in [0, 1]$ is the soiling factor of the PV system. The cleaner the surface of the PV system is, the larger c_{SF} is. The plane of the array irradiance E_{POA} and the PV cell temperature T_C are two dominant weather features. The calculation methods of E_{POA} and T_C are elaborated in the following sections.

1) Plane of Array Irradiance E_{POA}: E_{POA} is comprised of three components: i) beam component E_b, ii) ground-reflected component E_g and iii) sky-diffuse component E_d, shown as follows:

$$E_{POA} = E_b + E_g + E_d. \tag{2}$$

E_b is the direct irradiance from the sun onto the PV array, related to the direct normal irradiance (DNI) and the angle of incidence θ_{AOI}, shown as follows:

$$E_b = E_{DNI} \cos(\theta_{AOI}), \tag{3}$$

where θ_{AOI} is the angle between the sun rays and the PV array, determined by the hour angle, declination angle and zenith angle [23]. E_g is the irradiance on the PV array that is reflected off the ground, related to the global horizontal irradiance (GHI), expressed as follows:

$$E_g = c_{GR} E_{GHI}, \tag{4}$$

where c_{GR} is a constant factor, determined by the reflectivity of the ground and the tilt of the ground surface. E_d is a fraction of the diffuse irradiance from the sky dome, related to the diffuse horizontal irradiance (DHI). The isotropic sky model [24] of the DHI is expressed as follows:

$$E_d = c_{SKY} E_{DHI}, \tag{5}$$

where $c_{SKY} = 0.5(1 + \cos\theta_{T,array})$. In practice, E_{GHI} can be observed and forecasted from global satellite images. Then, given E_{GHI}, the values of E_{DNI} and E_{DHI} can be estimated based on the existing research [25]. In this paper, the empirical model in [26] is adopted to estimate E_{DNI} and E_{DHI} from E_{GHI}, simplified as follows:

$$E_{DNI} = f_{DNI}(E_{GHI}), E_{DHI} = f_{DHI}(E_{GHI}). \tag{6}$$

Therefore, E_{POA} can be derived from the three irradiance components:

$$E_{POA} = E_b + E_g + E_d = E_b + c_{GR} E_{GHI} + c_{SKY} E_{DHI}. \tag{7}$$

2) Photovoltaic Cell Temperature T_C: The PV cell temperature can be derived based on heat transfer with the ambient environment [27], shown as follows:

$$T_C = T_A + \frac{c_{TE} E_{POA}}{c_{W0} + c_{W1} V_W}, \tag{8}$$

where c_{TE} is a constant factor related to the adsorption efficiency of the PV system, c_{W0} and c_{W1} are the constant heat transfer factor and the convective heat factor, equaling 25 W/m^2·K and 6.84 W/m^3·s·K, respectively [27].

3) Reformulation: According to the models of E_{POA} and T_C, the block diagram of the calculation is shown in Fig. 2.

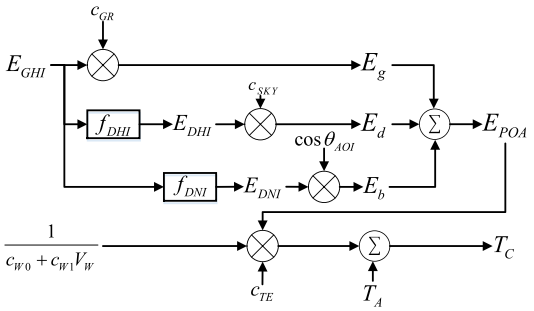

Fig. 2. The block diagram of the irradiance and temperature calculations.

In practice, some parameters are difficult to precisely acquire without field experiments, e.g., c_{GR} and c_{TE}. Therefore, according to (7) and (8), E_b, E_{GHI}, E_{DHI}, T_A and V_W are regarded as the variables instead of E_{POA} and T_C. Then, (1) can be reformulated as follows:

$$\begin{aligned} P_{mp} &= c_{STD} E_{POA} \cdot [1 + c_T (T_C - T_0)] \\ &= (k_1 + k_2 T_A)(E_b + k_3 E_{GHI} + k_4 E_{DHI}) \\ &\quad + k_5 \frac{(E_b + k_3 E_{GHI} + k_4 E_{DHI})^2}{c_{W0} + c_{W1} V_W}, \end{aligned} \tag{9}$$

where $k_i, i = 1, 2 \ldots 5$ are constant coefficients to be regressed. Given T historical observations, the matrix forms for E_b, E_{GHI}, E_{DHI} and T_A can be written as $\mathbf{E_b}, \mathbf{E_{GHI}}, \mathbf{E_{DHI}}, \mathbf{T_A} \in \mathbf{R}^{T\times 1}$, giving $\mathbf{E} = (\mathbf{E_b}, \mathbf{E_{GHI}}, \mathbf{E_{DHI}}) \in \mathbf{R}^{T\times 3}$. For simplification, $\mathbf{V_{W2}}$ is written as follows:

$$\begin{aligned} \mathbf{V_{W2}} = \Bigg(&\frac{1}{c_{W0} + c_{W1} V_W[1]}, \frac{1}{c_{W0} + c_{W1} V_W[2]}, \ldots, \\ &\frac{1}{c_{W0} + c_{W1} V_W[T]} \Bigg)^{\mathrm{T}} \end{aligned} \tag{10}$$

Thus, the key weather feature matrix $\mathbf{X_C}$ generated from the analytical modeling is:

$$\mathbf{X_C} = (\mathbf{E}, \mathrm{diag}(\mathbf{T_A}) \cdot \mathbf{E}, \mathrm{diag}(\mathbf{V_{W2}}) \cdot \mathbf{E_2}) \in \mathbf{R}^{T\times 12}, \tag{11}$$

where diag$(\cdot)$ is the function for diagonalization, and $\mathbf{E_2} \in \mathbf{R}^{T\times 6}$ is obtained by the expansion of E_{POA}^2 in (9):

$$\begin{aligned} \mathbf{E_2} = \Big(&E_b[t]^2, E_{GHI}[t]^2, E_{DHI}[t]^2, E_b[t] E_{GHI}[t], \\ &E_{GHI}[t] E_{DHI}[t], E_{DHI}[t] E_b[t] \Big), t = 1, 2, \ldots, T. \end{aligned} \tag{12}$$

Therefore, the key weather features, which are highly related to solar power, can be generated from the analytical model. In practice, other weather predictions may influence the accuracy of solar power forecast, e.g., humidity, pressure and precipitation. So the key weather feature matrix $\mathbf{X_C}$ and other weather predictions will be together taken as the input of machine learning methods. In contrast to existing forecasting methods, the key weather factors from the analytical modeling are used to reformulate the input of the machine learning methods, instead of directly inputting weather predictions.

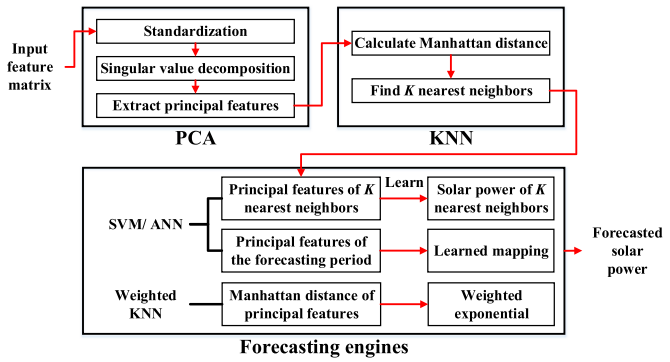

Fig. 3. The schematic of the machine learning methods.

C. Machine Learning Methods

In the machine learning module, PCA is used for feature extraction. KNN is employed to classify the forecasting period into the historical periods with the closest weather conditions. Then SVM, ANN and weighted KNN, are used as the forecasting engines. The schematic is shown in Fig. 3.

In Fig. 3, the input feature matrix $\mathbf{X_I} \in \mathbf{R}^{T\times N}$ is composed of key weather features $\mathbf{X_C}$ and other available weather predictions. The row of $\mathbf{X_I}$ represents each historical period and each column represents each feature. After PCA, the historical and forecasting principal component matrixes $\mathbf{X_P} \in \mathbf{R}^{T\times L}$ and $\mathbf{X_{PF}} \in \mathbf{R}^{H\times L}$ are simultaneously extracted. For each forecasting period from 1 to H, K nearest historical neighbors of each row of $\mathbf{X_{PF}}$ are then acquired from $\mathbf{X_P}$. Then three forecasting engines are employed to learn the mapping between the principal features and solar power of these K nearest neighbors. The learned mapping will be used for forecasting solar power.

PCA is a powerful tool in dimensional reduction and denoising for highly correlated data [28]. Since redundant data and the noise influence the forecasting accuracy, PCA is used in this paper to extract the principal components of the weather features while eliminating redundancy.

The first step of PCA is to standardize each column of $\mathbf{X_I}$ in the following manner [29]:

$$\mathbf{X_S}[i,j] = \frac{\mathbf{X_I}[i,j] - \min(\mathbf{X_I}[\cdot,j])}{\max(\mathbf{X_I}[\cdot,j]) - \min(\mathbf{X_I}[\cdot,j])}. \tag{13}$$

Then, the singular value decomposition (SVD) is applied to extract the principal components $\mathbf{X_P}$ and $\mathbf{X_{PF}}$.

KNN is aimed at classifying a forecasting period into K historical periods with the closest weather conditions. This step is known as similar-day selection [14].

In existing literature, it has been widely verified that the relationships between weather conditions and solar power highly depend on weather types [14], [30] shows the normalized root mean square errors (nRMSE) of solar power forecast with 15-min intervals are 7.85% for sunny days, 9.12% for cloudy days, 12.40% for rainy days and 12.60% for foggy days. Therefore, it is important to classify a forecasting period into a group of historical periods with closest weather conditions. Manhattan distance MD is selected as the measurement between forecasting and historical instances [31]:

$$MD(\mathbf{X_{PF}}[i,\cdot], \mathbf{X_P}[j,\cdot]) = \sum_{n=1}^{L} |\mathbf{X_{PF}}[i,n] - \mathbf{X_P}[j,n]|. \tag{14}$$

For the forecasting periods from 1 to H, K historical instances with smallest MD, will be selected from $\mathbf{X_P}$. It is worth mentioning that the value of K in KNN has an impact on the forecasting accuracy. If the value of K is too small, the training dataset will be insufficient due to the small sample size; If the value of K is too large, irrelevant historical periods will be selected into the training dataset, increasing forecasting errors. Thus, sensitivity analysis is conducted to determine the value of K during the cross-validation for the historical datasets.

In this paper, three well-known forecasting engines are adopted [14]. SVM is a promising tool to classify and regress with even small-sized training datasets [32]. The SVM regression performs linear regressions in the high-dimension feature space using ε–insensitive loss while reducing model complexity [33]. The kernel function is selected as the radial basis function (RBF). The hyperparameters in SVM, e.g., the penalty parameter in the objective and the deviation tolerance, are determined by sensitivity analysis.

Back propagation (BP) neural network is one of the most widely used ANNs. BP-ANN employs a multilayered feed-forward topology, and the basic structure is a three-layer network with the input, hidden and output layers [34].

In this paper, KNN is used to find K nearest neighbors, and the weighted KNN is used to calculate the forecasted solar power by an average weighted by exponential function [31]:

$$\hat{\mathbf{P}}_{\mathbf{mp}}[t] = \frac{\sum_{i=1}^{K} e^{-MD(\mathbf{X_{PF}}[t,\cdot],\mathbf{X_P}[i,\cdot])}\mathbf{P_{mp}}[i]}{\sum_{i=1}^{K} e^{-MD(\mathbf{X_{PF}}[t,\cdot],\mathbf{X_P}[i,\cdot])}}. \tag{15}$$

D. Deviation Analysis

In practice, the control scheme of a PV system is aimed at controlling its inverter to work in maximum power point tracking (MPPT) mode. However, it is difficult to instantly track the maximum power point on rainy, foggy and cloudy days [35]. These volatile weather conditions can greatly influence the performance of the inverters in PV systems. Thus, system bias of solar power forecast is inevitable in the real world. To reduce the errors caused by the system bias, a deviation analysis based on the cross-validation for historical datasets is conducted. A compensation term is then generated to adjust the forecasted solar power $\hat{\mathbf{P}}_{\mathbf{mp}}$.

The procedures of the deviation analysis for the forecasting period t are shown in Procedure 1.

For forecasting period t, the solar power outputted from the forecasting engine is $\hat{\mathbf{P}}_{\mathbf{mp}}[t]$. Then, the statistical deviations of the forecasted power within the neighbor $[\hat{\mathbf{P}}_{\mathbf{mp}}[t]\text{-}\varepsilon, \hat{\mathbf{P}}_{\mathbf{mp}}[t]\text{+}\varepsilon]$ can be obtained from the deviation distribution, where ε is a pre-defined constant. The compensation term $\Delta\mathbf{P}_{\mathbf{mp}}[t]$ is defined as the average of the forecasted deviations. The solar

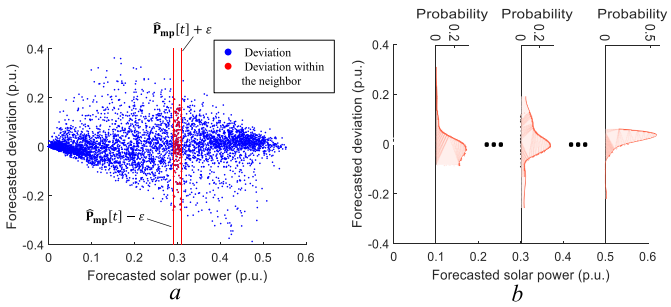

Fig. 4. The deviation distribution based on measured datasets from Australia.

Procedure 1 Deviation Analysis for Forecasting Period t

1 Randomly divide the K nearest neighbors into S equal-sized groups;
2 $i = 1$;
3 **Repeat**
4 Take the ith group as a validation;
5 Take the remaining S-1 groups for training;
6 Learn the mapping of the S-1 groups between the principal components and the solar power;
7 Define the forecasted deviations as the difference between the actual and forecasted solar power; Calculate the forecasted deviations of the ith group;
8 $i = i + 1$;
9 **until** $i = S + 1$.

power forecast is adjusted as follows:

$$\hat{\mathbf{P}}^{*}_{\mathbf{mp}}[t] = \hat{\mathbf{P}}_{\mathbf{mp}}[t] + \Delta\mathbf{P}_{\mathbf{mp}}[t]. \tag{16}$$

An example based on measured datasets from Australia is illustrated in Fig. 4. The forecasted deviations are shown in Fig. 4(a), and the probability distributions of the deviations are shown in Fig. 4(b) under different levels of forecasted solar power. As one can observe, the forecasting deviations obey a certain probability distribution, therefore it is reasonable to adjust the solar power forecast according to the statistical errors.

If the forecasted solar power from SVM is $\hat{\mathbf{P}}_{\mathbf{mp}}[t] = 0.3$, ε is set to 0.01 and the neighbor will be [0.3-0.01, 0.3+0.01]. Then, the compensation term $\Delta\mathbf{P}_{\mathbf{mp}}[t]$ is the average of the deviations within the neighborhood. Note that the value of ε has a significant impact on the value of $\Delta\mathbf{P}_{\mathbf{mp}}[t]$. If ε is too small, the points in the neighbor are sparse and the variation is large. If ε is too large, the forecasted solar power varies in a large neighbor and some points are not representative. Therefore, sensitivity analysis is conducted to determine the value of ε.

III. Case Studies

To verify the accuracy of the proposed forecasting approach, measured datasets from PV systems in Australia are used. Four error indices are adopted to measure the forecast accuracy [5]:

i) Normalized mean absolute error (nMAE)

$$\text{nMAE} = \frac{1}{H_2}\sum_{t=1}^{H_2}\frac{\left|\hat{\mathbf{P}}^{*}_{\mathbf{mp}}[t] - \mathbf{P}_{\mathbf{mp}}[t]\right|}{P_C} \times 100\%, \tag{17}$$

TABLE II
Forecasting Methods With and Without Key Weather Factors

No.	Forecasting Engine	Key weather factors	Other weather conditions
S1	SVM	√	√
S2	SVM	×	√
A1	ANN	√	√
A2	ANN	×	√
W1	Weighted KNN	√	√
W2	Weighted KNN	×	√
NR	Nonlinear regression	√	×

TABLE III
Description of Three PV Arrays in Australia

PV	Latitude	Longitude	P_C	$\theta_{A,array}$	$\theta_{T,array}$
1	35°16'30''S	149°6'49''E	1.56 kW	38°	36°
2	35°23'32''S	149°4'1''E	4.94 kW	327°	35°
3	35°32'S	149°9'E	4.00 kW	31°	21°

ii) Normalized root mean square error (nRMSE)

$$\text{nRMSE} = \frac{1}{P_C}\sqrt{\sum_{t=1}^{H_2}\frac{\left(\hat{\mathbf{P}}^{*}_{\mathbf{mp}}[t] - \mathbf{P}_{\mathbf{mp}}[t]\right)^2}{H_2}} \times 100\%, \tag{18}$$

iii) Normalized largest absolute error (nLAE)

$$\text{nLAE} = \frac{1}{P_C}\max\left\{\left|\hat{\mathbf{P}}^{*}_{\mathbf{mp}}[t] - \mathbf{P}_{\mathbf{mp}}[t]\right|, \forall t\right\} \times 100\%, \tag{19}$$

iv) Energy production error (EPE)

$$\text{EPE} = \frac{\left|\sum_{t=1}^{H_2}\left(\hat{\mathbf{P}}^{*}_{\mathbf{mp}}[t] - \mathbf{P}_{\mathbf{mp}}[t]\right)\right|}{\sum_{t=1}^{H_2}\mathbf{P}_{\mathbf{mp}}[t]} \times 100\%, \tag{20}$$

where H_2 is the number of forecasting periods, and the periods with 0 power are excluded. To demonstrate the effectiveness of the analytical modeling, different forecasting methods with and without the key weather factors are compared in the case studies. The forecasting methods are shown in TABLE II.

S1, A1 and W1 are the proposed methods using SVM, ANN and weighted KNN as the forecasting engine, respectively. By comparing different methods with and without key weather factors, the effects of the analytical modeling can be demonstrated. NR is a typical analytical method that uses nonlinear regression to estimate the parameters in the expression (9).

A. Solar Power Forecast Using Data in Australia

1) Data Description: In this section, the measured datasets from three PV arrays in Australia are tested [37]. The descriptions of the three PV arrays are shown in TABLE III.

The hourly historical weather predictions and solar power from April 1, 1:00, 2012 to June 30, 24:00, 2013 are taken as the training datasets. Seven scenarios are designed to test the forecasting performance, as shown in TABLE IV.

The weather data are from the datasets in [37]. The hourly weather predictions include 11 independent variables: total column liquid water, total column ice water, surface pressure, relative humidity, total cloud coverage, wind speed, temperature, GHI, surface thermal radiation, top net solar radiation E_a and total precipitation. The statistical descriptions of the historical datasets are shown in the Appendix.

TABLE IV
FORECASTING SCENARIOS

Scenario	Forecasting dataset
Spring	From October 1, 1:00, 2013 to October 31, 24:00, 2013
Summer	From January 1, 1:00, 2014 to January 31, 24:00, 2014
Autumn	From April 1, 1:00, 2014 to April 30, 24:00, 2014
Winter	From July 1, 1:00, 2013 to July 31, 24:00, 2013
Sunny	30 days with highest GHI from July 1, 2013 to May 1, 2014
Cloudy	30 days with highest cloud coverage from July 1, 2013 to May 1, 2014
Humid	30 days with highest relative humidity from July 1, 2013 to May 1, 2014

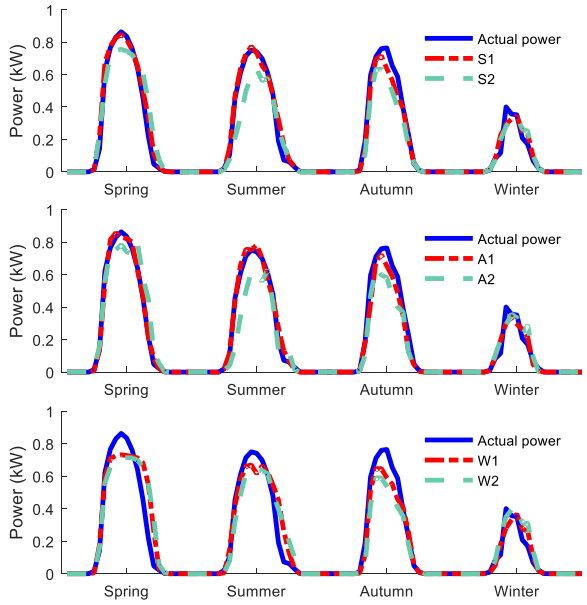

Fig. 5. Solar power of PV 1 on 18th of each month in four seasons using different methods.

TABLE V
FORECAST ERRORS IN SPRING SCENARIO IN AUSTRALIA (%)

PV	Index	S1	S2	A1	A2	W1	W2	NR
1	nMAE	4.03	6.44	**4.00**	6.43	6.63	7.37	7.89
	nRMSE	**6.26**	8.99	6.60	9.36	8.80	9.65	10.3
	nLAE	24.8	28.3	32.0	33.8	**24.0**	26.4	33.2
	EPE	3.09	5.00	**0.03**	3.17	2.40	4.01	18.9
2	nMAE	**1.37**	2.39	1.45	2.47	1.81	2.39	2.68
	nRMSE	**2.22**	3.17	2.24	3.27	2.50	3.08	3.32
	nLAE	10.5	9.50	9.18	9.55	**8.01**	8.77	9.57
	EPE	4.91	14.2	**1.02**	12.7	3.58	7.95	12.3
3	nMAE	**1.72**	2.47	1.92	2.51	2.67	2.92	2.74
	nRMSE	**2.77**	3.44	2.91	3.61	3.41	3.73	3.65
	nLAE	14.9	16.4	14.6	17.6	**13.9**	14.1	17.3
	EPE	0.67	3.33	0.94	2.90	**0.24**	1.90	10.5

As shown in (11), the key weather factors include 12 terms, which can be calculated based on the weather variables. The input features of the proposed methods, i.e., S1, A1 and W1, are the 11 available weather variables from the datasets and 12 key weather factors. The input features of S2, A2 and W2 are only the 11 available weather variables from the datasets. The input features of NR are only the 12 key weather factors. Then the forecasting is carried out day by day.

5-fold cross validation is conducted to determine the hyper-parameters in different forecasting methods [36], including the value of ε in the deviation analysis, the value of K in KNN, the parameters in SVM and the number of the hidden layer in ANN.

TABLE VI
FORECAST ERRORS IN SUMMER SCENARIO IN AUSTRALIA (%)

PV	Index	S1	S2	A1	A2	W1	W2	NR
1	nMAE	**3.02**	5.14	3.38	5.42	5.79	6.19	7.06
	nRMSE	**5.02**	7.25	5.64	7.92	7.81	8.08	9.19
	nLAE	26.1	25.6	26.8	30.8	**24.1**	32.2	24.7
	EPE	2.28	5.49	2.09	5.05	8.06	9.50	**0.79**
2	nMAE	**1.73**	2.63	1.75	2.69	2.22	2.75	3.05
	nRMSE	**2.48**	3.78	3.08	3.79	3.21	3.79	3.97
	nLAE	15.7	15.9	15.3	15.0	**13.0**	13.3	13.9
	EPE	**5.42**	17.0	4.16	16.0	10.9	17.3	16.5
3	nMAE	**1.41**	1.88	1.61	2.00	2.20	2.23	2.24
	nRMSE	**2.20**	2.63	2.36	2.84	3.05	3.05	2.98
	nLAE	12.4	10.9	12.2	13.7	**10.3**	10.6	11.7
	EPE	4.83	4.63	**1.40**	2.54	6.08	8.64	6.72

TABLE VII
FORECAST ERRORS IN AUTUMN SCENARIO IN AUSTRALIA (%)

PV	Index	S1	S2	A1	A2	W1	W2	NR
1	nMAE	**4.33**	5.05	4.45	5.08	5.31	5.57	5.57
	nRMSE	**6.52**	7.37	6.62	7.53	7.39	7.78	7.82
	nLAE	31.4	37.1	**28.8**	41.3	32.4	34.4	35.0
	EPE	2.24	9.40	4.65	10.5	**1.14**	11.2	7.63
2	nMAE	**1.50**	1.78	1.63	1.89	1.90	1.94	2.19
	nRMSE	**2.26**	2.51	2.34	2.61	2.63	2.64	2.93
	nLAE	9.36	8.61	8.72	9.04	**7.72**	9.21	9.67
	EPE	**1.01**	10.6	2.58	10.3	1.80	10.6	3.10
3	nMAE	**1.76**	2.01	1.96	2.21	2.19	2.27	2.11
	nRMSE	**2.83**	2.93	2.96	3.20	3.13	3.13	3.01
	nLAE	12.0	10.8	11.4	10.5	**9.90**	11.4.	11.7
	EPE	3.39	3.65	**1.14**	6.09	1.32	8.63	1.17

TABLE VIII
FORECAST ERRORS IN WINTER SCENARIO IN AUSTRALIA (%)

PV	Index	S1	S2	A1	A2	W1	W2	NR
1	nMAE	5.63	6.06	**5.51**	6.51	6.31	6.46	6.53
	nRMSE	8.66	9.03	**8.53**	9.59	9.21	9.31	8.94
	nLAE	32.6	31.2	33.7	33.8	33.0	31.9	**30.1**
	EPE	8.23	8.33	7.11	7.94	**4.24**	9.20	7.34
2	nMAE	**1.77**	1.81	1.84	1.84	2.05	2.12	2.17
	nRMSE	**2.50**	2.71	2.51	2.67	2.72	2.92	2.85
	nLAE	10.4	10.1	10.7	9.44	8.09	9.16	**8.07**
	EPE	**1.79**	10.6	2.66	9.92	10.6	9.31	2.42
3	nMAE	**2.04**	2.35	2.19	2.57	2.56	2.73	2.62
	nRMSE	**3.11**	3.31	3.14	3.54	3.54	3.75	3.51
	nLAE	13.3	12.2	12.0	12.6	13.0	12.1	**10.7**
	EPE	4.79	3.73	3.50	3.17	**0.53**	14.0	4.83

2) Forecast Results of Four Seasons: In this subsection, the forecast results of four seasons are compared with different forecasting methods. The actual and forecasted solar power of PV 1 on typical days in four seasons are compared in Fig. 5 using different methods.

The error indices of the forecast results are shown in TABLE V, TABLE VI, TABLE VII and TABLE VIII.

From the forecast results, a good match between the actual and forecasted solar power is obtained using the proposed methods. By comparing S1 with S2, A1 with A2 and W1 with W2 in four seasons, one can observe that the proposed methods outperform the traditional methods at nMAE and nRMSE due to taking key weather factors in account. For nLAE, the proposed methods with SVM, ANN or Weighted KNN

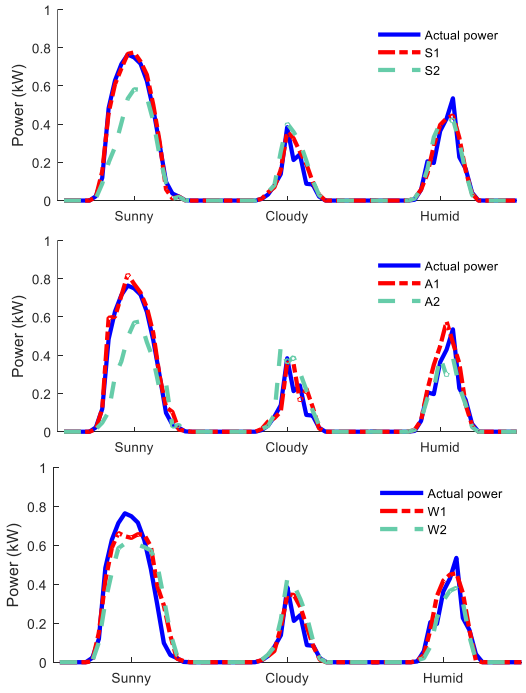

Fig. 6. Solar power of PV 1 on typical days of three weather types using different methods.

generally achieves the minimum except the cases in winter. As shown in TABLE VIII, NR has the minimum nLAE in winter while the proposed methods also have satisfactory performance. For EPE, the proposed methods have the minimum errors except for PV 1 in summer.

By comparing the forecast results in different seasons, one can observe that better forecast performance is achieved in summer while the winter generally has worse forecast performance. For nMAE, the minimum errors of three PV arrays are 3.02% (PV 1), 1.73% (PV 2) and 1.41% (PV 3) in summer while 5.51% (PV 1), 1.77% (PV 2) and 2.04% (PV 3) in winter. For nRMSE, the minimum errors of three PV arrays are 5.02% (PV 1), 2.48% (PV 2) and 2.20% (PV 3) in summer while 8.53% (PV 1), 2.50% (PV 2) and 3.11% (PV 3) in winter. This is because more sunny days are in summer than in winter.

3) Forecast Results of Three Weather Types: In this subsection, the forecast results of three weather types are compared with different forecasting methods. The actual and forecasted solar power of PV 1 on typical days of three weather are compared in Fig. 6 using different methods.

The error indices of the forecast results of three weather types are shown in TABLE IX, TABLE X and TABLE XI.

From the forecast results, the solar power can be forecasted in relatively high accuracy using the proposed methods. By comparing S1 with S2, A1 with A2 and W1 with W2 on sunny days, one can observe that the key weather factors greatly reduce the forecast errors of different methods. On cloudy and humid days, however, because the accuracy of the PV analytical modeling decreases, the key weather factors cannot be calculated precisely. Thus, the proposed methods have a good match between actual and forecasted power, but the performance is improved slightly by considering the key weather factors. For PV 1 and PV 2 on cloudy days, EPE might be increased by using the proposed methods. On humid days, S2 achieves the minimum nMAE of PV 2, and achieves the minimum nLAE of PV 3.

The average computation time of day-ahead forecasting using different forecasting methods in seven scenarios is shown in TABLE XII.

As one can observe, the proposed methods considering the calculation of key weather factors are efficient compared with the traditional methods. By comparing A1 with A2, the key

TABLE IX
FORECAST ERRORS ON SUNNY DAYS IN AUSTRALIA (%)

PV	Index	S1	S2	A1	A2	W1	W2	NR
1	nMAE	2.43	5.55	**2.35**	5.50	6.32	7.02	8.20
	nRMSE	**3.97**	8.14	4.09	8.45	8.86	9.33	10.6
	nLAE	**25.3**	28.2	29.1	30.6	30.5	35.2	32.7
	EPE	3.77	5.64	**0.87**	2.35	7.17	7.13	6.05
2	nMAE	**1.09**	2.47	1.27	2.48	1.71	2.60	2.77
	nRMSE	**2.20**	3.62	2.36	3.65	2.48	3.60	3.67
	nLAE	15.7	15.9	15.0	15.3	**13.1**	13.3	13.9
	EPE	4.85	17.9	**2.64**	16.0	4.57	12.6	13.4
3	nMAE	**0.87**	1.71	1.18	1.80	2.02	2.27	2.18
	nRMSE	**1.51**	2.52	1.80	2.63	2.70	3.11	3.08
	nLAE	20.5	20.6	19.3	19.3	**17.3**	17.8	22.4
	EPE	3.15	5.00	**0.11**	1.27	2.29	6.02	6.08

TABLE X
FORECAST ERRORS ON CLOUDY DAYS IN AUSTRALIA (%)

PV	Index	S1	S2	A1	A2	W1	W2	NR
1	nMAE	**4.07**	4.42	4.28	5.09	4.79	4.84	4.82
	nRMSE	**6.05**	6.56	6.30	7.42	6.86	7.06	7.10
	nLAE	23.8	24.1	28.2	24.8	**22.5**	25.8	30.9
	EPE	0.87	**0.84**	4.22	3.00	2.37	13.1	11.9
2	nMAE	**1.36**	1.37	1.44	1.51	1.53	1.54	1.72
	nRMSE	**2.00**	2.00	2.14	2.20	2.11	2.25	2.53
	nLAE	12.3	12.4	10.6	13.5	**9.82**	11.1	10.9
	EPE	1.60	7.14	0.63	5.64	10.0	**0.04**	22.2
3	nMAE	**1.61**	1.68	1.86	1.92	1.83	1.90	1.82
	nRMSE	**2.46**	2.55	2.63	2.72	2.61	2.77	2.75
	nLAE	11.2	13.0	**9.94**	10.2	10.8	12.6	15.1
	EPE	2.30	1.81	**1.12**	7.73	4.80	9.22	10.8

TABLE XI
FORECAST ERRORS ON HUMID DAYS IN AUSTRALIA (%)

PV	Index	S1	S2	A1	A2	W1	W2	NR
1	nMAE	**4.02**	4.05	4.20	4.80	4.30	4.78	4.89
	nRMSE	**5.81**	5.86	6.18	7.13	6.09	6.98	7.05
	nLAE	23.7	21.6	**21.5**	32.4	22.5	25.8	30.9
	EPE	16.8	21.0	12.6	25.9	**11.9**	31.4	33.8
2	nMAE	1.45	**1.37**	1.54	1.57	1.49	1.71	1.75
	nRMSE	2.20	2.20	2.29	2.34	**2.19**	2.52	2.55
	nLAE	12.3	12.4	12.2	11.1	**9.82**	11.1	10.9
	EPE	**2.40**	9.72	7.86	8.44	8.14	22.9	26.0
3	nMAE	**1.90**	1.98	2.04	2.22	2.06	2.12	2.24
	nRMSE	**3.00**	3.07	3.16	3.32	3.09	3.26	3.37
	nLAE	15.4	**12.3**	16.1	12.5	13.4	13.0	15.9
	EPE	9.97	13.3	12.3	19.1	**0.82**	20.5	24.7

TABLE XII
AVERAGE COMPUTATION TIME OF DAY-AHEAD FORECASTING

	S1	S2	A1	A2	W1	W2	NR
Time (s)	0.43	0.40	3.60	5.78	0.38	0.35	0.04

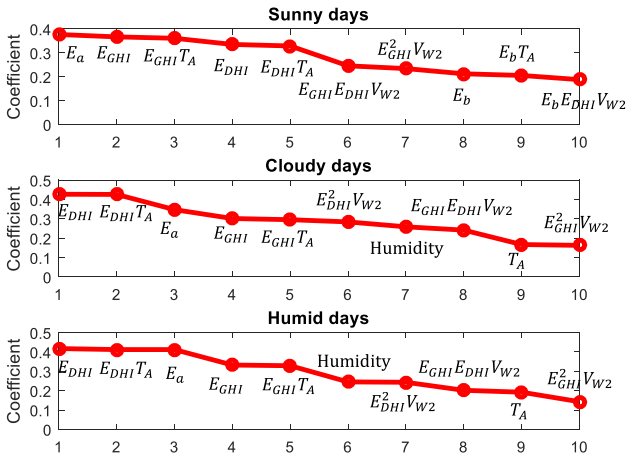

Fig. 7. Top 10 coefficients of the first column of $\mathbf{X_P}$ for PV 1 on sunny, cloudy and humid days.

weather factors can help ANN converge much faster. The average computation time is reduced by 37.72% after considering key weather factors.

4) Evaluation of Key Weather Factors: In this subsection, the contributions of the key weather factors are evaluated in different weather types. As shown in (21), the key weather factors include 12 terms:

$$\mathbf{X_C} = (\mathbf{E}, \mathrm{diag}(\mathbf{T_A}) \cdot \mathbf{E}, \mathrm{diag}(\mathbf{V_{W2}}) \cdot \mathbf{E_2}) \in \mathbf{R}^{T\times 12}, \quad (21)$$

which are three irradiance components and the combination of irradiance, ambient temperature and wind speed. To extract the principal components of weather features, PCA is used in this paper. The number of the extracted principal components is decided according to cumulative percentage of explained variance, which is set to 95%. Based on the singular value decomposition (SVD), the relationship between the input matrix $\mathbf{X_I}$ and the principal component $\mathbf{X_P}$ is as follows:

$$\begin{aligned} \mathbf{X_P} &= \mathbf{X_I} \cdot \mathbf{C}, \\ \mathbf{C}^{\mathrm{T}}\mathbf{C} &= \mathbf{I}, \mathbf{C} \in \mathbf{R}^{N\times L}, \end{aligned} \quad (22)$$

where $\mathbf{C} \in \mathbf{R}^{N\times L}$ is the coefficient matrix, the elements of which represent the contributions of $\mathbf{X_I}$ to $\mathbf{X_P}$.

For PV 1 on sunny, cloudy and humid days, the top 10 coefficients of the first column of $\mathbf{X_P}$ are shown in Fig. 7. As one can observe, the input weather features have different contributions and impacts on the principal components, and the key weather factors derived from the analytical modeling are significant. On sunny days, the top 10 factors are the combinations of different irradiance components, ambient temperature and wind speed. Top net solar radiation E_a and global horizontal irradiance E_{GHI} are the top 2 factors because of the clear sunny days. On cloudy and humid days, the top 2 factors are the diffuse horizontal irradiance E_{DHI} and its product with ambient temperature $E_{DHI}T_A$. This indicates that the sky-diffuse irradiance components greatly influence the solar power on cloudy and humid days. Note that humidity is another significant factor on cloudy and humid days.

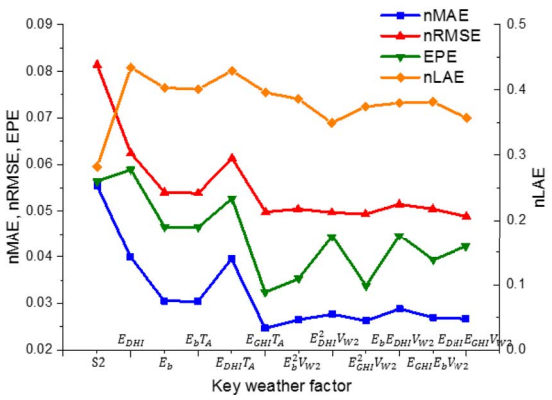

Fig. 8. Forecast errors of key weather factors in $\mathbf{X_C}$ for PV 1 on sunny days.

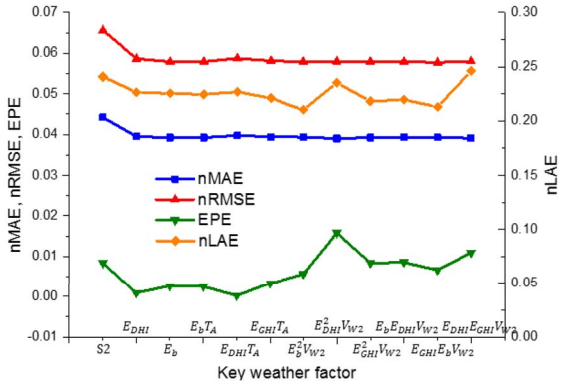

Fig. 9. Forecast errors of key weather factors in $\mathbf{X_C}$ for PV 1 on cloudy days.

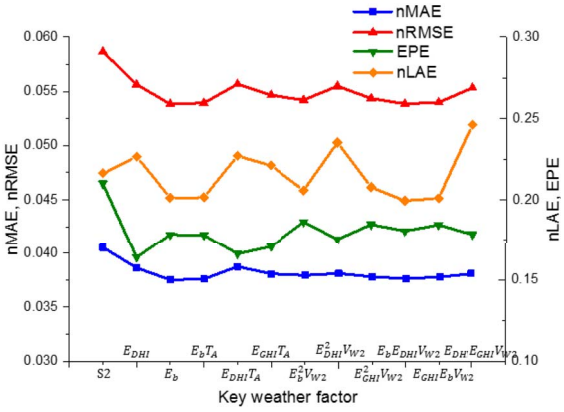

Fig. 10. Forecast errors of key weather factors in $\mathbf{X_C}$ for PV 1 on humid days.

To quantify the contributions of the key weather factors, the forecast performance is tested by adding each key weather factor at one time. Fig. 8, Fig. 9 and Fig. 10 show the forecast errors of each key weather factor in $\mathbf{X_C}$ compared with S2 on sunny, cloudy and humid days, respectively. Note that the global horizontal irradiance E_{GHI} is initially considered in S2.

From the forecast results, the impacts of each key weather factor can be obtained. After adding each key weather factor, nMAE and nRMSE are significantly reduced while nLAE and EPE vary irregularly. On sunny days, by adding

TABLE XIII
STATISTICAL DESCRIPTIONS OF THE HISTORICAL DATASETS OF PV 1

Weather feature	Mean	Median	Standard deviation
Total column liquid water (kg/m^3)	0.04	0	0.11
Total column ice water (kg/m^3)	0.02	0	0.05
Surface pressure (10^3 Pa)	94.67	94.68	0.58
Relative humidity (%)	57.82	56.87	19.45
Total cloud cover	0.43	0.36	0.40
Wind speed (m/s)	3.42	2.98	2.09
Temperature (K)	289.24	288.85	7.00
Global horizontal irradiance (J/m^2)	405.29	364.90	305.35
Surface thermal radiation (J/m^2)	317.77	317.74	41.34
Top solar radiation (J/m^2)	476.33	452.46	319.58
Total precipitation (10^{-5} m)	7.88	0	35.52
Solar power (kW)	0.33	0.26	0.28

TABLE XIV
STATISTICAL DESCRIPTIONS OF THE HISTORICAL DATASETS OF PV 2

Weather feature	Mean	Median	Standard deviation
Total column liquid water (kg/m^3)	0.04	0	0.12
Total column ice water (kg/m^3)	0.02	0	0.05
Surface pressure (10^3 Pa)	94.12	94.13	584.20
Relative humidity (%)	56.23	54.81	19.38
Total cloud cover	0.43	0.37	0.39
Wind speed (m/s)	3.17	2.74	1.95
Temperature (K)	289.00	288.60	6.84
Global horizontal irradiance (J/m^2)	409.57	370.47	302.68
Surface thermal radiation (J/m^2)	315.35	315.23	40.94
Top solar radiation (J/m^2)	483.14	461.86	317.42
Total precipitation (10^{-5} m)	7.97	0	34.98
Solar power (kW)	0.37	0.34	0.29

$E_{GHI}T_A$, nMAE and nRMSE both achieve the minimum. In addition, by adding the key factors related to E_{DHI} and E_b, the forecast performance can be improved on cloudy and humid days because the diffuse irradiance component is important.

IV. CONCLUSION

In this paper, a novel solar power forecasting approach is proposed by exploring the key weather factors from the analytical modeling of PV systems. The analytical model of a PV system can provide physical knowledge regarding solar power, and the machine learning methods explore the nonlinear and time-varying nature of a practical PV system. In the analytical modeling module, key weather features are theoretically derived from weather predictions. In contrast to existing methods, the input of the machine learning methods is reformulated using both key weather features and available weather predictions. To further reduce forecasting errors, deviation analysis for the historical datasets is conducted to adjust the forecasted solar power.

Case studies based on the measured datasets from PV systems in Australia demonstrate that by taking advantage of the key weather features derived from PV models, the forecasting performance can be effectively improved. In the future, two aspects deserve further investigations: i) The key weather factors derived from different PV analytical models and their contributions to solar power forecast; ii) The forecast of the ramping events in solar power caused by sudden weather changes. The proposed approach will possibly provide new insights for solar power forecasting technique.

TABLE XV
STATISTICAL DESCRIPTIONS OF THE HISTORICAL DATASETS OF PV 3

Weather feature	Mean	Median	Standard deviation
Total column liquid water (kg/m^3)	0.04	0	0.12
Total column ice water (kg/m^3)	0.02	0	0.05
Surface pressure (10^3 Pa)	92.47	92.49	575.63
Relative humidity (%)	57.06	55.48	19.94
Total cloud cover	0.44	0.40	0.39
Wind speed (m/s)	3.13	2.74	1.85
Temperature (K)	288.06	287.69	6.72
Global horizontal irradiance (J/m^2)	401.88	360.55	303.91
Surface thermal radiation (J/m^2)	307.36	306.31	40.58
Top solar radiation (J/m^2)	475.38	452.71	320.68
Total precipitation (10^{-5} m)	8.60	0	36.57
Solar power (kW)	0.38	0.35	0.30

The detailed data can be found in [37].

APPENDIX

The statistical descriptions of the historical datasets are shown in TABLE XIII, TABLE XIV and TABLE XV. Note that the periods with zero solar power output are excluded.

The detailed data can be found in [37].

REFERENCES

[1] Y. Ding, C. Kang, J. Wang, Y. Chen, and B. F. Hobbs, "Foreword for the special section on power system planning and operation towards a low-carbon economy," *IEEE Trans. Power Syst.*, vol. 30, no. 2, pp. 1015–1016, Mar. 2015.

[2] J. Wang, H. Zhong, Z. Ma, Q. Xia, and C. Kang, "Review and prospect of integrated demand response in the multi-energy system," *Appl. Energy*, vol. 202, pp. 772–782, Sep. 2017.

[3] *Mercom Forecasts 76 GW in Global Solar Installations in 2016, A 48% Increase Over 2015*, Mercom Capital Group, Austin, TX, USA, 2016. [Online]. Available: http://mercomcapital.com/mercom-forecasts-76-gw-in-global-solar-installations-in-2016-a-48-yoy-increase-over-2015

[4] H.-T. Yang, C.-M. Huang, Y.-C. Huang, and Y.-S. Pai, "A weather-based hybrid method for 1-day ahead hourly forecasting of PV power output," *IEEE Trans. Sustain. Energy*, vol. 5, no. 3, pp. 917–926, Jul. 2014.

[5] E. B. Ssekulima, M. B. Anwar, A. A. Hinai, and M. S. El Moursi, "Wind speed and solar irradiance forecasting techniques for enhanced renewable energy integration with the grid: A review," *IET Renew. Power Gener.*, vol. 10, no. 7, pp. 885–989, Jun. 2016.

[6] Z. Li *et al.*, "Short-term wind power prediction based on extreme learning machine with error correction," *Protect. Control Modern Power Syst.*, vol. 1, no. 1, pp. 1–8, 2016.

[7] F. Almonacid, P. J. Pérez-Higueras, E. F. Fernández, and L. Hontoria, "A methodology based on dynamic artificial neural network for short-term forecasting of the power output of a PV generator," *Energy Convers. Manag.*, vol. 85, pp. 389–398, Sep. 2014.

[8] H. Tian, F. Mancilla-David, K. Ellis, E. Muljadi, and P. Jenkins, "A cell-to-module-to-array detailed model for photovoltaic panels," *Solar Energy*, vol. 86, no. 9, pp. 2695–2706, 2012.

[9] H. S. Sahu and S. K. Nayak, "Numerical approach to estimate the maximum power point of a photovoltaic array," *IET Gener. Transm. Distrib.*, vol. 10, no. 11, pp. 2670–2680, Aug. 2016.

[10] A. S. B. M. Shah, H. Yokoyama, and N. Kakimoto, "High-precision forecasting model of solar irradiance based on grid point value data analysis for an efficient photovoltaic system," *IEEE Trans. Sustain. Energy*, vol. 6, no. 2, pp. 474–481, Apr. 2015.

[11] A. J. Veldhuis *et al.*, "An empirical model for rack-mounted PV module temperatures for southeast Asian locations evaluated for minute time scales," *IEEE J. Photovolt.*, vol. 5, no. 3, pp. 774–782, May 2015.

[12] C. Yang, A. A. Thatte, and L. Xie, "Multitime-scale data-driven spatio-temporal forecast of photovoltaic generation," *IEEE Trans. Sustain. Energy*, vol. 6, no. 1, pp. 104–112, Jan. 2015.

[13] C. Wan *et al.*, "Photovoltaic and solar power forecasting for smart grid energy management," *CSEE J. Power Energy Syst.*, vol. 1, no. 4, pp. 38–46, Dec. 2015.

[14] Y. Zhang, M. Beaudin, R. Taheri, H. Zareipour, and D. Wood, "Day-ahead power output forecasting for small-scale solar photovoltaic electricity generators," *IEEE Trans. Smart Grid*, vol. 6, no. 5, pp. 2253–2262, Sep. 2015.

[15] H. S. Jang, K. Y. Bae, H.–S. Park, and D. K. Sung, "Solar power prediction based on satellite images and support vector machine," *IEEE Trans. Sustain. Energy*, vol. 7, no. 3, pp. 1255–1263, Jul. 2016.

[16] M. J. Sanjari and H. B. Gooi, "Probabilistic forecast of PV power generation based on higher order Markov chain," *IEEE Trans. Power Syst.*, vol. 32, no. 4, pp. 2942–2952, Jul. 2017.

[17] H. Sheng, J. Xiao, Y. Cheng, Q. Ni, and S. Wang, "Short-term solar power forecasting based on weighted Gaussian process regression," *IEEE Trans. Ind. Electron.*, to be published.

[18] J. Liu, W. Fang, X. Zhang, and C. Yang, "An improved photovoltaic power forecasting model with the assistance of aerosol index data," *IEEE Trans. Sustain. Energy*, vol. 6, no. 2, pp. 434–442, Apr. 2015.

[19] J. R. Andrade and R. J. Bessa, "Improving renewable energy forecasting with a grid of numerical weather predictions," *IEEE Trans. Sustain. Energy*, vol. 8, no. 4, pp. 1571–1580, Oct. 2017.

[20] A. Tascikaraoglu *et al.*, "Compressive spatio-temporal forecasting of meteorological quantities and photovoltaic power," *IEEE Trans. Sustain. Energy*, vol. 7, no. 3, pp. 1295–1305, Jul. 2016.

[21] J. M. Filipe, R. J. Bessa, J. Sumaili, R. Tomé, and J. N. Sousa, "A hybrid short-term solar power forecasting tool," in *Proc. 18th Int. Conf. Intell. Syst. Appl. Power Syst.*, Porto, Portugal, 2015, pp. 1–6.

[22] B. Marion, "Comparison of predictive models for photovoltaic module performance," in *Proc. 33rd IEEE Photovolt. Specialists Conf.*, San Diego, CA, USA, 2008, pp. 1–6.

[23] I. Reda and A. Andreas, "Solar position algorithm for solar radiation applications," *Solar Energy*, vol. 76, no. 5, pp. 577–589, 2004.

[24] P. G. Loutzenhiser *et al.*, "Empirical validation of models to compute solar irradiance on inclined surfaces for building energy simulation," *Solar Energy*, vol. 81, no. 2, pp. 254–267, 2007.

[25] F. J. Batlles, M. A. Rubio, J. Tovar, F. J. Olmo, and L. Alados-Arboledas, "Empirical modeling of hourly direct irradiance by means of hourly global irradiance," *Energy*, vol. 25, no. 7, pp. 675–688, 2000.

[26] D. T. Reindl, W. A. Beckman, and J. A. Duffle, "Diffuse fraction correlations," *Solar Energy*, vol. 45, no. 1, pp. 1–7, 1990.

[27] D. Fairman, "Assessing the outdoor operating temperature of photovoltaic modules," *Progr. Photovolt. Res. Appl.*, vol. 16, no. 4, pp. 307–315, 2008.

[28] K. Benidis, Y. Sun, P. Babu, and D. P. Palomar, "Orthogonal sparse PCA and covariance estimation via procrustes reformulation," *IEEE Trans. Signal Process.*, vol. 64, no. 23, pp. 6211–6226, Dec. 2016.

[29] I. B. Mohamad and D. Usman, "Standardization and its effects on K-means clustering algorithm," *Res. J. Appl. Sci. Eng. Technol.*, vol. 6, no. 17, pp. 3299–3303, 2013.

[30] J. Shi, W.-J. Lee, Y. Liu, Y. Yang, and P. Wang, "Forecasting power output of photovoltaic systems based on weather classification and support vector machines," *IEEE Trans. Ind. Appl.*, vol. 48, no. 3, pp. 1064–1069, May/Jun. 2012.

[31] Y. Zhang and J. Wang, "GEFCom2014 probabilistic solar power forecasting based on K-nearest neighbor and kernel density estimator," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, Denver, CO, USA, 2015, pp. 1–5.

[32] Y. Wang, Q. Xia, and C. Kang, "Secondary forecasting based on deviation analysis for short-term load forecasting," *IEEE Trans. Power Syst.*, vol. 26, no. 2, pp. 500–507, May 2011.

[33] P.-H. Chen, C.-J. Lin, and B. Schölkopf, "A tutorial on V-support vector machine," *Appl. Stochastic Models Bus. Ind.*, vol. 21, no. 2, pp. 111–136, 2005.

[34] M. Cui, D. Ke, D. Gan, and Y. Sun, "Statistical scenarios forecasting method for wind power ramp events using modified neural networks," *J. Mod. Power Syst. Clean Energy*, vol. 3, no. 3, pp. 371–380, 2015.

[35] N. Mutoh, M. Ohno, and T. Inoue, "A method for MPPT control while searching for parameters corresponding to weather conditions for PV generation systems," *IEEE Trans. Ind. Electron.*, vol. 53, no. 4, pp. 1055–1065, Jun. 2006.

[36] J. Friedman, T. Hastie, and R. Tibshirani, *The Elements of Statistical Learning*. New York, NY, USA: Springer, 2001.

[37] Global Energy Forecasting Competition. (2014). *IEEE Working Group on Energy Forecasting*. [Online]. Available: http://www.gefcom.orgl

Jianxiao Wang (S'14) received the B.S. degree in electrical engineering from Tsinghua University, Beijing, China, in 2014, where he is currently pursuing the Ph.D. degree. He is a visiting student researcher with Stanford University, Stanford, CA, USA. His research interests include power system operation and planning, and electricity market.

Haiwang Zhong (S'10–M'13) received the Ph.D. degree in engineering from Tsinghua University, Beijing, China, in 2013, where he is currently working as an Assistant Professor. His research interests include generation scheduling optimization, demand response, and electricity markets.

Xiaowen Lai received the B.E. and Ph.D. degrees in electrical engineering from Tsinghua University, Beijing, China, in 2010 and 2015, respectively, where he is currently a Post-Doctoral Fellow. His research interest includes electricity market, economic dispatch, and demand response.

Qing Xia (M'01–SM'08) received the Ph.D. degree in engineering from Tsinghua University, Beijing, China, in 1989, where he is currently a Professor. His research interests include electricity markets, generation scheduling optimization, and power system planning.

Yang Wang received the B.E. and Ph.D. degrees in electrical engineering from Tsinghua University, Beijing, China, in 2006 and 2011, respectively, where he is currently a Scientist. His research interest includes electricity market, economic dispatch, and demand response.

Chongqing Kang (M'01–SM'07–F'17) received the Ph.D. degree in engineering from Tsinghua University, Beijing, China, in 1997, where he is currently a Professor. His research interests include power system planning, power system operation, renewable energy, low carbon electricity technology, and load forecasting.