IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 28, 2019, accepted January 9, 2020, date of publication January 27, 2020, date of current version February 5, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2969549

Solar Irradiance Capturing in Cloudy Sky Days–A Convolutional Neural Network Based Image Regression Approach

HUAIGUANG JIANG, (Member, IEEE), YI GU, YU XIE, RUI YANG, (Member, IEEE), AND YINGCHEN ZHANG, (Senior Member, IEEE)
National Renewable Energy Laboratory, Golden, CO 80210, USA

Corresponding author: Yingchen Zhang

This work was supported in part by the National Renewable Energy Laboratory (NREL), operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract under Grant DE-AC36-08GO28308, and in part by the Laboratory Directed Research and Development (LDRD) Program, NREL.

ABSTRACT Global horizontal irradiance (GHI) is a critical index to indicate the output power of the photovaltaic (PV). In traditional approaches, the local GHI can be measured with very expensive instruments, and the large-area GHI collection depends on complex satellite-based models, solargis algorithms, and the high-performance computers (HPC). In this paper, a novel approach is proposed to capture the GHI conveniently and accurately. Considering the nonstationary property of the GHI on cloudy days, the GHI capturing is cast as an image regression problem. In traditional approaches, the image regression problem is treated as two parts, feature extraction (for the images) and regression model (for the regression targets), which are optimized separately and blocked the interconnections. Considering the nonlinear regression capability, a convolutional neural network (CNN) based image regression approach is proposed to provide an End-to-End solution for the cloudy day GHI capturing problem in this paper. The multilayer CNN is based on the AlexNet and VGG. The L2 (least square errors) with regularization is used as the loss function in the regression layer. For data cleaning, the Gaussian mixture model with Bayesian inference is employed to detect and eliminate the anomaly data in a nonparametric manner. The purified data are used as input data for the proposed image regression approach. In the experiments, three-month sky images and GHI data (with 1-min resolution) are provided by the National Renewable Energy Laboratory (NREL) with the HPC system. The numerical results demonstrate the feasibility and effectiveness of the proposed approach.

INDEX TERMS Convolutional neural network, solar irradiation, global horizontal irradiance, image regression, variational inference, Bayesian theory, Dirichlet process, deep learning, sky image.

NOMENCLATURE

PARAMETERS AND VARIABLES

$t_{i_2}^{sr}$: The times of sunrise for day i_2
$t_{i_2}^{ss}$: The times of sunset for day i_2
$\beta_{miss}(t_2)$: The missing GHI data at time t_2
h_{j_2}, a_{j_2}: Coefficients of the spline interpolation
X: The sample set or the observation set (GHI)
Z: The component index set
π: The mixing weights
μ: The mean of the normal distributions
Λ: The precision matrix
$p(Z|\pi)$: Given mixing weights, the conditional distribution of Z
$p(X|Z, \mu, \Lambda)$: Given latent variables Z, μ, Λ, the conditional distribution of $p(X|Z, \mu, \Lambda)$
$Dir(\pi|\alpha_0)$: The Dirichlet distribution with given concentration parameter α_0
$C(\alpha_0)$: The normalization item for Dirichlet distribution
$p(\mu, \Lambda)$: The joint probability distribution of μ and Λ
Θ: A sample space with a distribution G_0
$\mathcal{N}(\cdot)$: The normal distribution
$Beta(1, \beta)$: Beta distribution with stick-breaking construction and parameters 1 and β

The associate editor coordinating the review of this manuscript and approving it for publication was Hui Liu.

$q_t(\cdot)$: The defined exponential distribution family
$D(q_t(Z)||p(Z|X,\vartheta))$: The KL divergence between $q_t(Z)$ and $p(Z|X,\vartheta)$
$\mathcal{X}$: The collected sky image set
$\mathcal{Y}$: The collected GHI value set
D: The training samples $D = \{\mathbf{x}_i, y_i\}_{i=1}^{N_1}$
E: The objective function with mean-square-error
δ^m: The error in last layer m
w_{jl}^{k+1}: The weight for node j for incoming node l in layer $k+1$
o_i^{k-1}: The output for node i in layer k
H_1, W_1: The hight and width of the sky image
k_1, k_2: The hight and width of the filter in 2-dimensional CNN
b_{i_1,j_1}^k: The bias for node with position i_1, j_1 at layer k in 2-dimensional CNN
χ_{i_1,j_1}^k: The convolved input vector at layer k with position i_1, j_1 in 2-dimensional CNN

I. INTRODUCTION

With a low cost, the penetration of rooftop PV is increasing fast in recently years, which can be seemed as an indispensable component for modern power system integration [1]–[5]. Solar power, especially behind-meter solar power forecast is an important aspect for operating the grid with high level of PV penetration [6], [7]. The GHI is defined as the sum of the total received direct and diffuse shortwave solar irradiation on a horizontal plane. It is widely used to indicate the output power of PV [6], [7]. In traditional approaches, the local GHI can be collected with very expensive instruments, which are inconvenient to be moved and can hardly be used for large-area monitoring. For large-area monitoring, the radiative transfer models are widely used. However, these depend on the complex models and HPC. In this paper, a deep learning-based approach, a CNN-based approach is proposed to capture the GHI conveniently and accurately, which can also be implemented into large-area monitoring.

As shown in Fig. 1, in traditional approach (a), GHI is routinely measured by ground-based radiometers, e.g. pyranometers, use either thermoelectric or photoelectric detectors [8]. The reliability of this measurements is highly depend on installation scheme, hardware maintenance, calibration technology and frequency [9]. The availability of the measurements is often restricted by the high cost. In traditional approach (b), GHI can be numerically simulated by radiative transfer models, which account for the atmospheric absorption and scattering by air molecules, aerosols, and clouds [10]. However, accurate simulation of GHI replies on comprehensive measurements or retrievals of atmospheric constituents and land surface that are sparsely available over large scale areas. Conventional radiative transfer models are often challenged by computation complexity by solving the radiative transfer equation and considering the interactions between the atmosphere and land surface. In traditional approach (c), the processing approach includes two parts: feature engineering and further processing [11]–[14]. Feature engineering is similar as feature extraction, which extract the image features with different approaches such as filtering, Fourier transform, principle component analysis, wavelet analysis, and autoencoder related approaches [12], [15]. The extracted features are the inputs for next step, which is used for classification, pattern recognition, anomaly detection, and regression with different approaches such as hidden Markov model, support vector machine, neural network, logistical regression, random forest, and Bayesian network [12], [16]–[20]. The two parts framework in approach (c) is disconnected and has no communication between them, which results in a lower the performance of the approach [12], [21], [22]. In the approach (d), it provides an End-to-End learning framework for the image regression task, which combines the feature extraction and regression part together to higher the efficiency and performance. As discussed above, the Pros and Cons of the 4 methods are summarized and presented in Table 1.

In this paper, thee-month data (minutes level resolution data) are provided by the NREL. Considering the huge volume of data, a data cleaning procedure is used to eliminate the errors and inconsistencies for improving the input data quality, which contains data selection, calibration, missing data reconstruction, data standardization, and normalization. This procedure is significantly indispensable and greatly impacts the final performance. Gaussian mixture model is an useful method to detect the anomaly data, however, the number of the mixture components is difficult to determine. Based on the Dirichlet process, variational inference, and Bayesian theory, a Gaussian mixture model with Bayesian inference approach is employed to determine the number of the mixture components automatically [23]–[29]. Compared with Markov Chain Monte Carlo (MCMC), the variational inference is fast (it doesn't require many samples and saves sampling time). The expectation propagation is very similar with the variational inference, which can be regarded as the same category [23]–[25], [35]. Considering the real-application in power systems, the important factors such as temperature and Zenith angle are collected to detect the anomaly data.

In cloudy days, the profile of solar irradiation has a lot of stochastic deviations such as abrupt decreasing and increasing, which cause the deviations of output power of the PV. With the increasing penetration of PV, the large deviations of the PV output power result in a series problem such as voltage deviation, frequency oscillation, even unplanned islanding. Based on this [30]–[34], the proposed CNN based image regression model is focusing on building a relationship between the input cloudy sky images and the solar irradiations. Recently, the CNN based image processing, especially for image classification, is developing very fast. In 2012 ImageNet Large Scale Visual Recognition Competition (ILSVRC), the AlexNet (a type of CNN) [37] is proposed with 16.4 % error rate, which is considerably better than the previous shallow machine learning model (similar as approach (c) in Fig.1 with error 25.8 %). From then on, the CNN based image processing is attracting more attention

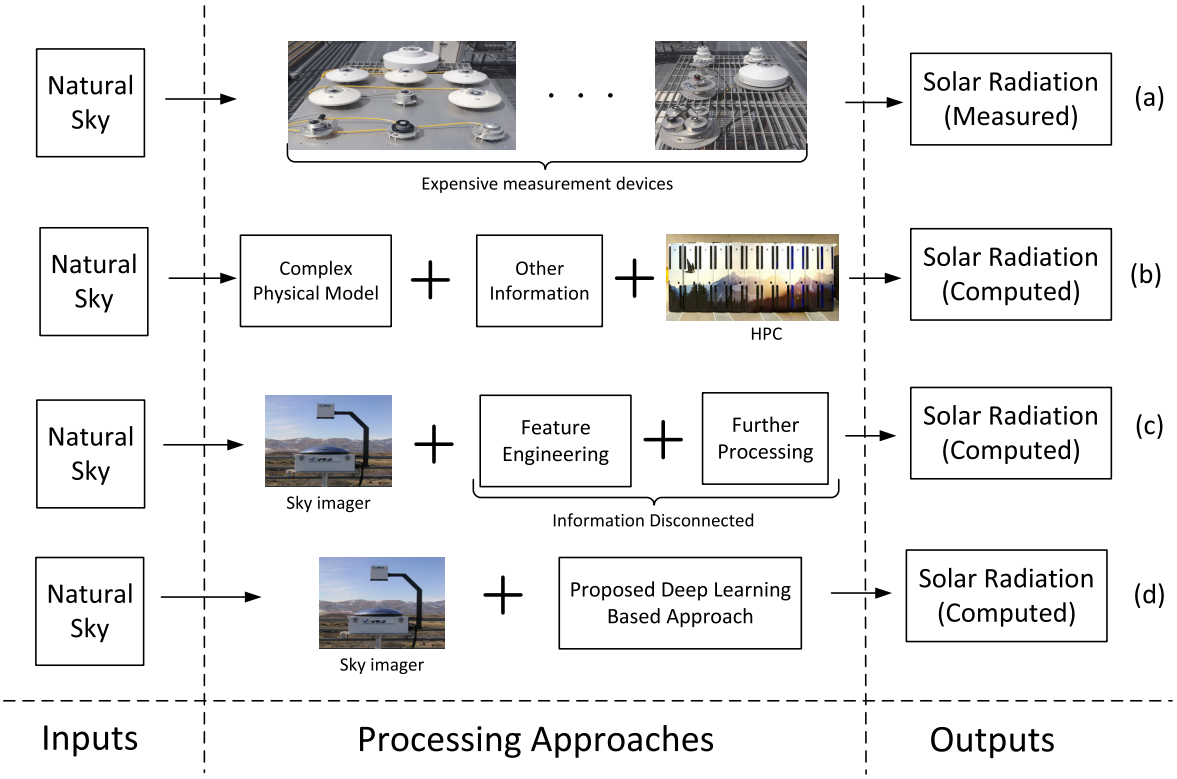


FIGURE 1. The proposed approach compared with the traditional approaches.

TABLE 1. The comparison of different approaches in Fig.1.

Approaches	Pros	Cons
(a)	The solar irradiation can be measured fast and accurately	Expensive device, inconvenient to move, calibration needed, weather fragile (broken by hails) and limited observation area
(b)	large-area solar irradiation can be computed	Complex physical model and atmosphere model, satellite image and HPC are needed
(c)	Can be extended to large-area monitoring	Low efficiency for parameter turning and low performance, large training data needed
(d)	Fast, accurately, cheap, and can be extended to large-area monitoring conveniently	large training data needed

in computer vision area. In 2014, the VGG with 19 layers and GoogleNet with 22 layer achieved 7.3 % and 6.7 % error rate, respectively, which further demonstrate the capability of CNN in computer vision processing [38], [39]. In 2015, a big improvement is provide by the 152 layers ResNet with 3.57 % error rate, which is better than human expert (5.1 % error rate) [40].

The proposed image regression problem is similar with the age estimation problem in computer vision area, which want to build a map between the input human images and the corresponding ages. In [41], the age estimation is studied with a shallow machine learning model, support vector machine. In [42], a CNN based regression approach is proposed to estimate human age with the images of human face as the input data. However, the proposed multi-output CNN is also a classification problem, and people with different ages are classified into different small groups. Because the solar irradiation is a continuous variable with a relatively big range, the classification of the solar irradiation brings a big error and deviation for the system. Therefore, in this paper, a CNN based image regression approach is proposed to provide a fast and accurate solar irradiation capturing.

The **main contributions** of this paper are:

1) A big data processing approach is used in renewable energy area for solar irradiation capturing, which contains big data cleaning and deep learning based image regression. Compared with the traditional solar irradiation capturing approach, the CNN based approach is cheap, fast, accurate, and convenient to be extended for large-area monitoring.
2) Based on the Dirichlet process, variational inference, and Bayesian theory, a Gaussian mixture model with Bayesian inference approach is employed to determine the mixture components automatically. Considering the real-application in power systems, the important factors such as temperature and Zenith angle data are also collected to detect and eliminate the anomaly data.
3) Based on the CNN architecture for classification, a new regression CNN architecture is designed for image regression problem. The input of the proposed approach is the image set and the output is the continuous variable set, which can be extended for multiple regression problems. According to Deep dream related algorithms, the CNN working manner is illustrated layer by layer with Figures, which help researchers deeply understand the working manner of deep learning. In the future study, the related problems, for example, the large-area PV output forecasting, can be studied based on our research.

As shown in the Fig. 2, this paper consists of three major components: data cleaning, CNN based image regression, and results validation, which are described as follows. In Section II, data cleaning is introduced with 7 steps, and the Bayesian inference is used to detect the anomaly data in a nonparametric manner. In Section III, the CNN based image regression approach is introduced for solar irradiation capturing, and the characteristic of the proposed CNN architecture is detailed analyzed. In Section IV, the numerical results are presented to validate the proposed approach. Finally, the conclusion is presented in Section V.

FIGURE 2. The flowchart of proposed approach.

II. DATA CLEANING

In machine learning, data cleaning is a significant component, which tremendously affect the performance of the whole approach. In this paper, the original data including the GHI data and sky images are directly collected from the devices in NREL's Solar Radiation Research Laboratory, which needs the data cleaning for the CNN learning. As shown in Fig. 3, the proposed data cleaning approach contains 7 steps: data merge, data calibration, missing data recovery, data normalization, anomaly detection, data standard, and data verify and map.

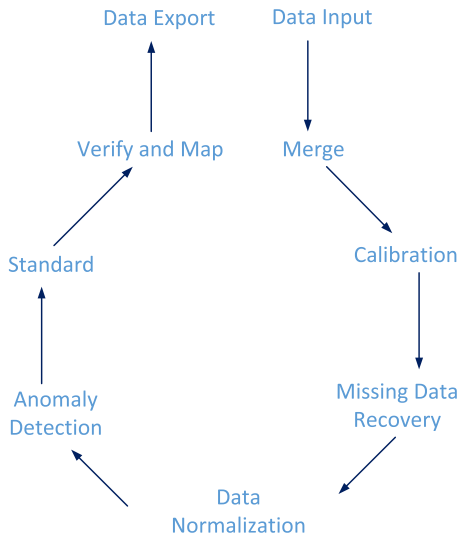


FIGURE 3. The flowchart of data cleaning in proposed approach.

A. DATA MERGE, CALIBRATION, AND MISSING DATA RECOVERY

In this paper, the data merge, calibration, and missing data recovery are explained together as pre-data processing for the data cleaning. The original data including the GHI data and sky image data are collected with sample rate 1 sample/min. Because the period of the sky image is 24 hours, we merge 24 hours data as a section for each day. Because the CNN is used to build a regression relationship between the sky images and GHI data, the original data need to be calibrated and recovered, and the problem is list as follows:

1) The sky image camera only capturing the sky image from sunrise to sunset, the length is variant everyday (with some random offsets of the camera).
2) The GHI data are collected 24 hours, but the corresponding time index is Greenwich Mean Time (GMT).
3) There are random sky images and GHI data missing in the sky image series and GHI data everyday, which cause the mismatching between the sky images and GHI data.

The times of sunrise $t_{i_2}^{sr}$ and sunset $t_{i_2}^{ss}$ for day i_2 are collected from National Oceanic and Atmosphere Administration (NOAA). The accurate length between $t_{i_2}^{sr}$ and $t_{i_2}^{ss}$ can be calculated as: $\Delta t_{i_2} = t_{i_2}^{ss} - t_{i_2}^{sr}$, which can be used to eliminate the random offsets of the camera, and delete the useless sky image captured in the night (as shown in Fig. 5). Then, the corresponding GHI data can be sectioned with Δt_{i_2}, and the time index is transfered from GMT to Mountain Time (MT).

In the collected original data, there are random sky images and GHI missing. For the sky images missing, the corresponding GHI data are deleted to keep the one to one mapping relationship. Because the sky images is continuously changing and the sample rate 1 sample/min is relatively high, the missing GHI data can be recovered by the spline interpolation as equation 1. $\beta_{miss}(t_2)$ is the missing GHI data, t_{2,j_2}

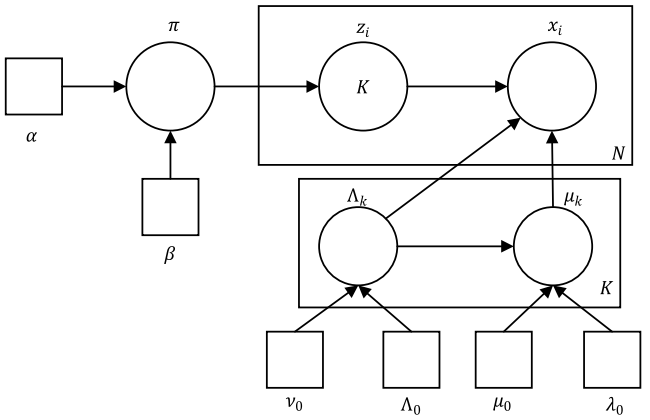


FIGURE 4. The graphic model of the Gaussian mixture model with Bayesian inference.

indicates the time index j_2 to discriminate with the missing GHI time t_2, h_{j_2} and a_{j_2} are coefficients. Then, the missing GHI data at time t_2 can be recovered.

B. ANOMALY DETECTION

1) PROBLEM DESCRIPTION

In this paper, we collected a large volume of original data from the sky image camera and GHI sensors, which contains a lot of anomaly data sets. These anomaly data sets can pollute the input data, and generate irrelevant information during the learning process [43], [44], [46]–[48]. As in Fig.5, because we are focusing on the solar irradiance capturing for the cloudy sky days (causing large deviations of PV output power), this means that the sky images of the sunrise, sunset, and clean days are seemed as the anomaly data and need to be cleaned. (1), is obtained by the equation as shown at the bottom of this page.

In oder to detect the anomaly conveniently, the data are normalized to range (0, 1) with the approach in [13], and the corresponding sky images are stored and tagged in a database. In this paper, a nonparameter and fast anomaly detection method is employed to detect the anomaly data as shown in Fig. 5.

2) PROBLEM FORMULATION

As shown in Fig. 4, the anomaly detection method is based on Gaussian mixture model with Bayesian inference [23]–[27]. The full joint distribution can be formulated as follows:

$$p(X, Z, \pi, \mu, \Lambda) = p(X|Z, \mu, \Lambda)p(Z|\pi)p(\pi)p(\mu|\Lambda)p(\Lambda) \tag{2}$$

where

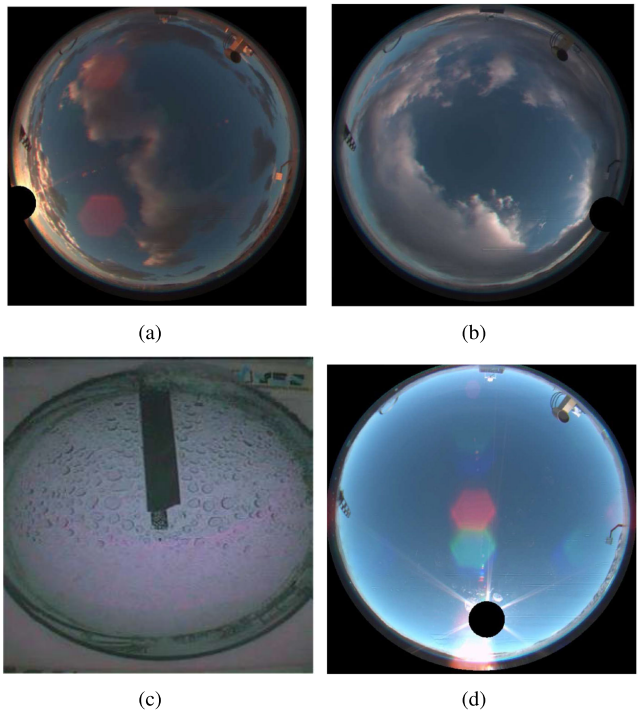


FIGURE 5. The selected anomaly sky images taken by different devices (a) the sky image of sunrise, (b) the sky image of sunset, (c) the sky image polluted by rain and snow, (d) the sky image of a clean day without any clouds.

1) X is the sample set or the observation set, $X = \{x_1, \cdots, x_N\}$. In this paper, X is the collected GHI.
2) Z is the component index set $Z = \{z_1, \cdots, z_N\}$, which also indicate the total category is K.
3) π is the mixing weights, specifically, different from Dirichlet distribution, in Dirichlet process $\pi = \{\pi_1, \pi_2, \cdots, \pi_\infty\}$.
4) μ is the mean of the normal distributions, which is with a normal distribution prior as shown in Fig. 4 with a mean μ_0 and precision matrix λ_0.
5) Λ is the precision matrix, which is with a Wishart prior (or Gamma, Gamma distribution is equivalent to a one-dimensional Wishart distribution [49]) with a scale matrix Λ_0 and a degree of freedom ν_0.

In the Gaussian mixture model (GMM), the number of the Gaussian components gives the prior distribution estimation of the whole distribution. If this number is given, the whole distribution can be solved in a convenient manner. Therefore, the character Z here is used as a latent variable to indicate the number of the Gaussian components, which is formulated as

$$\beta_{miss}(t_2) = \frac{a_{j_2+1}(t_2 - t_{2,j_2})^3 + a_{j_2}(t_{2,j_2+1} - t_2)^3}{6h_{j_2}} + \left(\frac{\beta_{j_2+1}}{h_{j_2}} - \frac{h_{j_2}}{6}a_{i_2+1}\right)(t_2 - t_{j_2}) + \left(\frac{\beta_{j_2}}{h_{j_2}} - \frac{h_{j_2}}{6}a_{i_2}\right)(t_{j_2+1} - t_2) \tag{1a}$$

$$h_{j_2} = t_{2,j_2+1} - t_{2,j_2}, \quad \text{and } a_0 = 0, \text{ and } a_{n_2} = 0 \tag{1b}$$

$$h_{i_2-1}a_{j_2-1} + 2a_{j_2}(h_{i_2-1} + h_{i_2}) + a_{j_2+1}h_{j_2} = 6\left(\frac{\beta_{j_2+1} - \beta_{j_2}}{h_{j_2}} - \frac{\beta_{j_2} - \beta_{j_2-1}}{h_{j_2-1}}\right) \tag{1c}$$

a categorical distribution and its conjugate distribution is the Dirichlet distribution.

Based on (2), the relationship and probabilities can be formulated as follows. The conditional distribution of Z given a mixing weight:

$$p(Z|\pi) = \prod_{n_3=1}^{N_3}\prod_{k_3=1}^{K_3}\pi_k^{z_{nk}} \tag{3}$$

Then, given the latent variables Z, μ, Λ, the conditional distribution of the observed data can be formulated as:

$$p(X|Z,\mu,\Lambda) = \prod_{n_3=1}^{N_3}\prod_{k_3=1}^{K_3}\mathcal{N}(x_{n_3}|\mu_{k_3},\Lambda_{k_3}^{-1})^{Z_{n_3,k_3}} \tag{4}$$

Here, the Bayesian inference is employed to estimate how many clusters are required for the observation data, which means the hyperparameters can be generated with noninformative hyperprior distributions. Firstly, the nonparametic prior titled as Dirichlet distribution is employed to build the finite Gaussians. Then, the Dirichlet process is used to generalize it into infinite Gaussians.

$$p(\pi) = Dir(\pi|\alpha_0) = C(\alpha_0)\prod_{k_3=1}^{K_3}\pi_{k_3}^{\alpha_0-1} \tag{5}$$

where Dir is the Dirichlet distribution, α_0 is the concentration parameter, and $C(\alpha_0)$ is a normalizing item which can be expressed with Gamma Function and also be named as multivariate Beta function [24], [49], [50]. Then, according to (2), the Gaussian-Wishart prior can be introduced for the mean μ and precision matrix Λ.

$$p(\mu,\Lambda) = p(\mu|\Lambda)p(\Lambda) \tag{6a}$$

$$= \prod_{k_3=1}^{K_3}\mathcal{N}(\mu_{k_3}|\mu_0,(\lambda_0\Lambda_{k_3})^{-1})\mathcal{W}(\Lambda_{k_3}|\Lambda_0,\nu_0) \tag{6b}$$

where μ_0 is a mean, λ_0 is a precision matrix, Λ_0 is a scale matrix, and ν_0 is degree of freedom.

As in (5), the Dirichlet distribution can be seemed as the conjugate prior for the categorical distribution. To generalize it into infinite and nonparametric distribution, the Dirichlet process can be seemed as its conjugate prior. For a sample space Θ, G_0 is a distribution over Θ with a positive factor β, and the Dirichlet process can be generated with G_0 over Θ.

$$(G(A_1),\cdots,G(A_{K_3})) \sim Dir(\beta G_0(A_1),\cdots,\beta G_0(A_{K_3})) \tag{7a}$$

$$G \sim DP(\beta, G_0) \tag{7b}$$

where A_{K_3} is a finite measurable partition over Θ, and the positive factor β controls the density of G [24], [26]. Here, the stick-breaking construction is used for the weights π_{k_3}:

$$\pi'_{k_3} \sim Beta(1,\beta) \tag{8a}$$

$$\pi_{k_3} = \pi'_{k_3}\prod_{i_3=1}^{k_3-1}(1-\pi'_{i_3}) \tag{8b}$$

where $\pi = \{\pi_1,\cdots,\pi_\infty\}$ which indicate the original model is generalized into infinite Gaussians with Dirichlet process.

3) VARIATIONAL INFERENCE

In this paper, the Bayesian theorem based variational inference algorithm is employed for the Dirichlet process based Gaussian mixture models. The posterior distribution of Z can be computed as

$$\log p(Z|X,\vartheta) = \log p(X,Z|\vartheta) - \log p(X|\vartheta) \tag{9}$$

where ϑ is the parameters, and this Bayesian theorem based variational inference provide a bridge to the likelihood function and prior function, which can also be seemed as a regularization item for the likelihood function.

Here, we introduce $q_\iota(W)$, $q_\iota(W)$ is defined as a distribution family, for example, the exponential distribution family, and ι is the parameter. Then, according to Jensen's equation:

$$\log p(X|\vartheta) \geq E_q[\log p(Z,X|\vartheta)] - E_q[\log_{q_\iota}(Z)] \tag{10}$$

where (10) is the evidence of low bound (ELOB), and the gap is the Kullback-Leibler (KL) divergence between $q_\iota(Z)$ and $p(Z|X,\vartheta)$, which can be derived as

$$D(q_\iota(Z)||p(Z|X,\vartheta)) = E_q[\log_{q_\iota}(Z)] - E_q[\log p(Z,X|\vartheta)] + \log p(X|\vartheta) \tag{11}$$

In optimization, we can maximize the ELOB, which is an alternatively option for minimizing (11) [51], [52]. And the detail information can be found in [23]–[27].

Specifically, the temperature is an indispensable factor to impact the efficiency of the PV, PV battery, and related systems [53]. In this paper, we also collect corresponding temperature information to eliminate the anomaly data as shown in Fig. 5. Similarly, the Zenith angle is a critical factor for the GHI [54], which is also collected to eliminate the anomaly data.

After this process, the anomaly data shown in Fig. 5 can be detected and only the sky images with clouds can be selected for next step. Then, all the selected sky images are merged together with the temporal and GHI information for CNN regression.

III. CONSTITUTIONAL NEURAL NETWORK BASED IMAGE REGRESSION

A. PROBLEM FORMULATION

After the data cleaning as discussed above, the collected sky images and GHI values are formed as one-to-one correspondence. Then, an input space $\mathcal{X} = \{\mathbf{x}_1,\cdots,\mathbf{x}_{N_1}\}$ consists of the collected sky images. Because a sky image is a three dimensional matrix including red, green, and blue color, we use $\mathbf{x}_i$ to denote the i-th sky image. The corresponding output space $\mathcal{Y} = \{y_1,\cdots,y_{N_1}\}$ consists of the collected GHI values, where y_i is the i-th GHI value. Given the training samples $D = \{\mathbf{x}_i, y_i\}_{i=1}^{N_1}$, the proposed CNN based image regression approach aims to find a mapping from images to GHI values $h(\cdot)$: $\mathcal{X} \mapsto \mathcal{Y}$ with a predefined cost function C: $\mathcal{X}\times\mathcal{Y} \mapsto R$.

From the traditional optimization perspective, in the training part, the cost function $C(h)$ needs to be minimized with several different approaches such as stochastic gradient

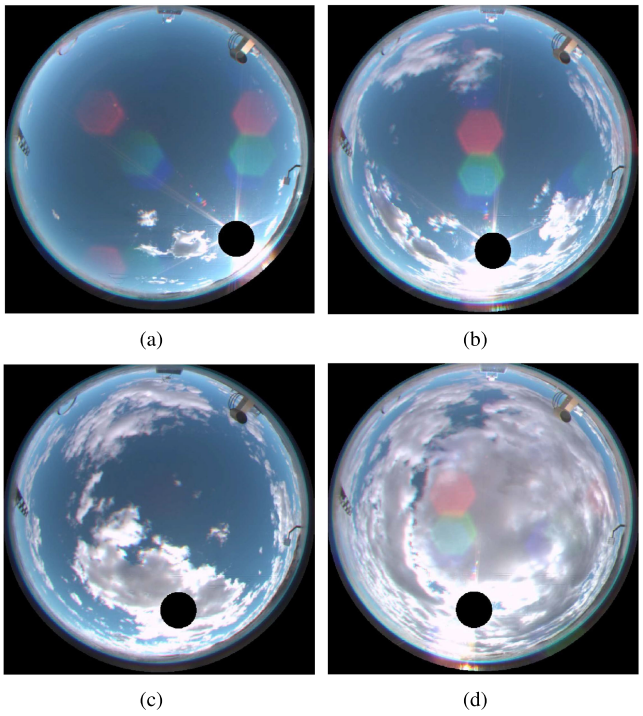


FIGURE 6. The selected normal sky images with some clouds (a) few cloud with okta 1 to 2 in a sky image, (b) cloud with okta 3 to 4 in a sky image, (c) cloud with okta 5 to 6 in a sky image, (d) cloud with okta 7 to 8 in a sky image.

descent (SGD), momentum, and Nesterov momentum, and in this paper, ADAM is used [22], [55]. However, different from the traditional optimization, in machine learning, the distribution of the training space is usually different from the distribution in the testing space, which requires a good generalization characteristic for the selected $h(\cdot)$. If using the traditional optimization approaches without any revision during the training, several problems such as ill-condition, local minimum, cliffs, and etc. can dramatically damage the testing performance. Therefore, the are several basic methods such as early stopping, parameters/hyperparameters initialization, adaptive learning, and the detailed information can be found in [22].

B. THE ARCHITECTURE OF THE PROPOSED CNN

After data clean, the sky images of normal cloudy days are selected as shown in Fig. 6, where okta is a unit of measurement to describe the cloud cover in meteorology [56]. Compared with the images in ImageNet (a large-scale image database) [57], as shown in Fig. 6, the patterns of the cloud in the sky images are simpler, which doesn't requires very complex architectures.

As shown in Fig. 7, the architecture of the proposed CNN is designed with 5 convolution layers, batch normalization, rectified linear unit (ReLU) (activation function), and max-pooling, which is based on the VGG 16 architecture [38], [39]. For each convolutional block, the detailed design is shown in Fig. 7 and described as follows:

1) The first part is the convolutional layer. Compared with Alexnet $11 * 11$ or other $7 * 7$ perception fields, $3 * 3$ is implemented to increasing the nonlinear characteristics and reduce the computation load [38], [39]. At the same time, the stride step is set as 1 and padding is also set as 1, which keeps the size of the sky image for next block processing [22], [38], [39].
2) The second part is the batch normalization part, which is aiming to solve the interal covariance shift [45]. With the increasing layers of the CNN, the distribution of the inputs changes gradually, which causes the proposed CNN unstable and very difficult to train. With the batch normalization part, the unwanted distribution shift can be reduced, the training speed can be increased, and the proposed CNN is more stable [45]. And the inference part of the batch normalization can be found in [22], [45].
3) The third part is the ReLU part, which can be seemed as the activation function and brings nonlinear characteristics for the proposed CNN. Compared with other activation functions such as sigmoid, hyberbolic tangenet, and Gaussian, the ReLU function is more convenient to compute the derivative, fewer vanishing gradient, and fewer saturate parts [22]. Specifically, in this proposed CNN, the target is the solar irraidance, which means all the output should be positive real numbers. The ReLU function can meet this requirement perfectly [22].
4) The parameter set for the convolutional block, for example (112, 112, 48), indicates the output dimension of the convolutional block is $112*112$, and 48 indicates there are 48 feature maps (or feature images). This means the features can be selected in different feature maps, which provide a more convenient way for the feature extraction.
5) The last part is the max pooling part. In this paper, the max pooling is design as $2*2$ with stride $2*2$, which select the maximum value over the $2 * 2$ part. This part can be seemed as a down-sampling part, which select the feature of the $2*2$ and non-overlapping area, reduce the dimensionality of the input, and pass the selected features to the next convolution block.

After 5 convolutional block, the dimensional of the input image is changed from $224 * 224 * 3 = 150,528$ to $7 * 7 * 388 = 19,012$, which means only 12.63% data can represent the features of the original image. Then, the output is flatten into a vector and treat as an input for the fully connected layer with 1000 neurons for the final regression.

C. LEARNING OF THE PROPOSED CNN

1) LAST LAYER

In classification problem, the last layer usually uses the softmax activation function and cross entropy as objective function. However, considering the regression objective, the mean-square-error is employed as the objective function,

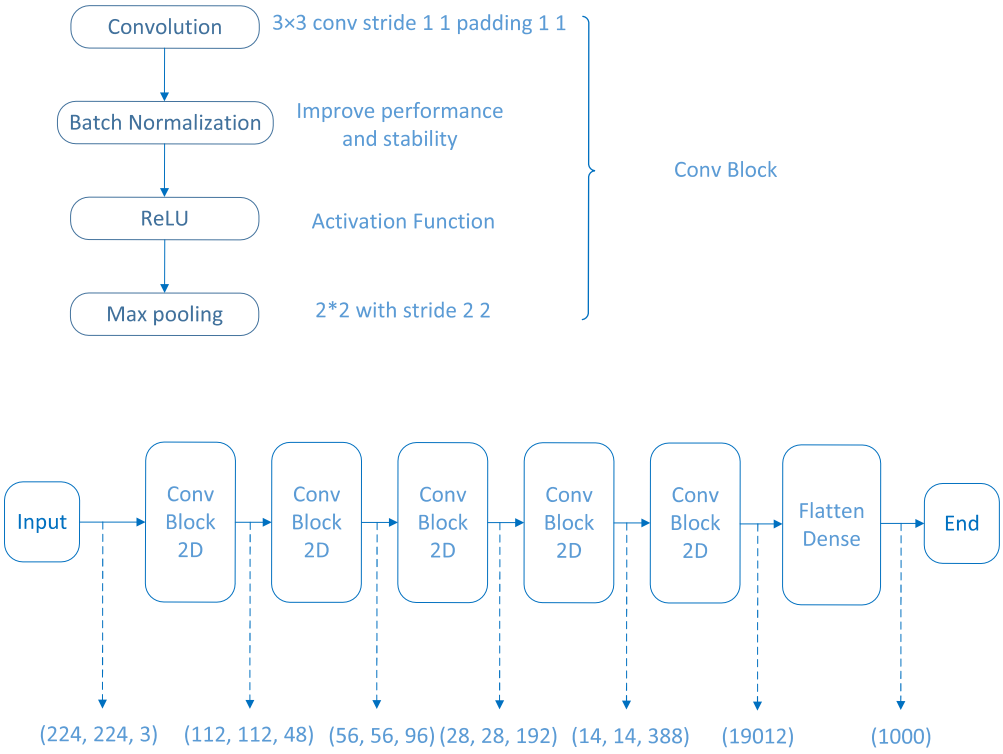


FIGURE 7. The architecture of the proposed CNN.

and linear function is used as activation function.

$$E = \frac{1}{2}(g_{final}(\chi_1^m) - y)^2 + reg_{l2} \tag{12}$$

where g_{final} is the activation function, which is ReLU as shown in Fig. 7, χ_1^m is the only one output in last layer m, y is a general form of the GHI value, and reg_{l2} is a weighted L2 regularization item, and the elastic (combined L1 and L2) can also be implemented here [24], and E is the loss. Then, the error can be derived as:

$$\delta_1^m \equiv \frac{\partial E}{\partial \chi_1^m} = (g_{final}(\chi_1^m) - y) g'_{final}(\chi_1^m) \tag{13}$$

where the derivative of ReLU is

$$g'_{final}(\chi_1^m) = \begin{cases} 1, & if\ \chi_1^m > 0 \\ 0, & otherwise \end{cases} \tag{14}$$

2) HIDDEN LAYER

In the hidden layer, the error item can be derived as following:

$$\delta_j^k = g'(\chi_j^k) \sum_{l=1} w_{jl}^{k+1} \delta_l^{k+1} \tag{15}$$

where $k+1$ indicates the layer, δ^k is the error in last layer k, w_{jl}^{k+1} is a weight for node j for incoming node l in layer $k+1$. And the weight can be derived as:

$$\frac{\partial E}{\partial w_{ij}^k} = o_i^{k-1} g'(\chi_j^k) \sum_{l=1}^{r^{k+1}} w_{jl}^{k+1} \delta_l^{k+1} \tag{16}$$

where o_i^{k-1} is the output for node i in layer k, r is the total number for node l.

3) EXTENDED TO 2-DIMENSIONAL CNN

In this paper, the proposed approach is based on CNN and aims to build a mapping between sky images and GHI values, which requires a extension from 1 dimensional backward propagation to 2 dimensional backward propagation. The weight can be computed as

$$\frac{\partial E}{\partial w_{m',n'}^k} = \sum_{i_1=0}^{H_1-k_1} \sum_{j_1=0}^{W_1-k_2} \delta_{i_1,j_1}^k o_{i_1+m',j_1+n'}^{k-1} = \widehat{\delta_{i_1,j_1}^k} * o_{m',n'}^{k-1} \tag{17}$$

where $*$ indicates a 2-dimensional (H_1 and W_1 indicates the hight and width of the sky image) convolution in CNN [22], [55], $k_1 \times k_2$ is the dimension of the filter, $\widehat{\delta_{i,j}^k}$ indicates the flipped kernel [58]. δ and o are the same as defined above. Then, similar with equation (13) the error can be computed as:

$$\frac{\partial E}{\partial \chi_{i',j'}^k} = \sum_{m=0}^{k_1-1} \sum_{n=0}^{k_2-1} \delta_{i'-m,j'-n}^{k+1} w_{m,n}^{k+1} f'\left(\chi_{i',j'}^k\right) \tag{18}$$

where χ is defined as:

$$\chi_{i_1,j_1}^k = \sum_m \sum_n w_{m,n}^l o_{i_1+m,j_1+n}^{k-1} + b_{i_1,j_1}^k \tag{19}$$

where b_{i_1,j_1}^k is the bias for node with position i_1, j_1 in layer k. In sum, the 2-dimensional CNN formulated for the proposed sky image processing.

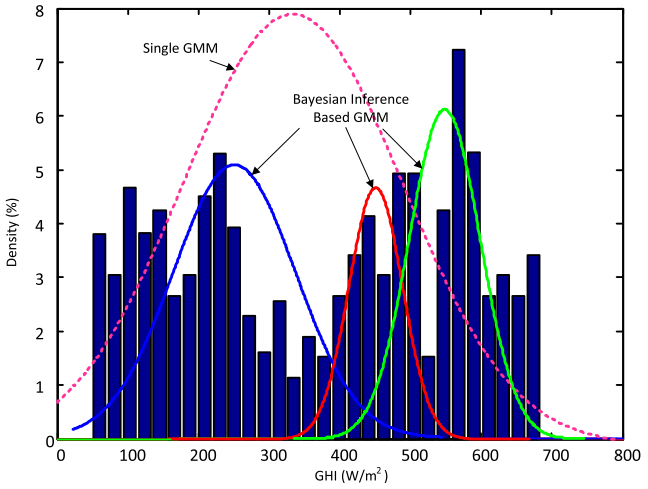


FIGURE 8. Anomaly detection for the GHI data.

IV. NUMERICAL RESULTS

In this paper, the image data and GHI data are provided NREL, which includes about thee-month data in the Autumn of 2016 and Summer of 2017. The sampling rate of the sky imager is 1 sample/min from sunrise to sunset per day. The sampling rate of the GHI measurement device is also 1 sample/min for 24 hours per day.

A. ANOMALY DETECTION AND ELIMINATION

First, the original GHI data are modeled with the Gaussian mixture model with Bayesian inference. Compared with the Gaussian mixture model without Bayesian inference (indicated as Single GMM with magenta), the GHI data can be more precisely modeled with 3 Gaussian models instead of 1 Gaussian model. According to our requirements, the GHI values located in the areas close to 0 and 800 can be considered as sunrise, sunset, and clean sky data.

Second, as mentioned above, the Zenith angle and temperature are considered as useful factors, the Gaussian mixture model with Bayesian inference are employed to detect the anomaly data with them. To show the results clearly, the result of Zenith angle and temperate are shown separately as in Fig. 9 and Fig. 10.

In the upper part of Fig. 9, the collected data can be modeled with a single Gaussian model, which means a lot of useful data are classified into anomaly data side. In the lower part, the collected data can be modeled with 4 Gaussian models, the anomaly data are indicated corresponding to the sky images of sunrise and sunset (with the green dots close to 0 GHI and 50° Zenith angle). The yellow dots located close to 800 GHI are corresponding to the sky images of clean sky without clouds.

Similarly, in the upper part of Fig. 10, the collected data can be modeled with 2 single Gaussian model, which means a lot of useful data are classified into anomaly data side. In the lower part, the collected data can be modeled with 5 Gaussian models, the anomaly data are indicated corresponding to the sky images of sunrise and sunset (with the yellow dots close to 0 GHI and 10°C temperature). The light blue dots located close to 800 GHI are corresponding to the sky images of clean sky without clouds.

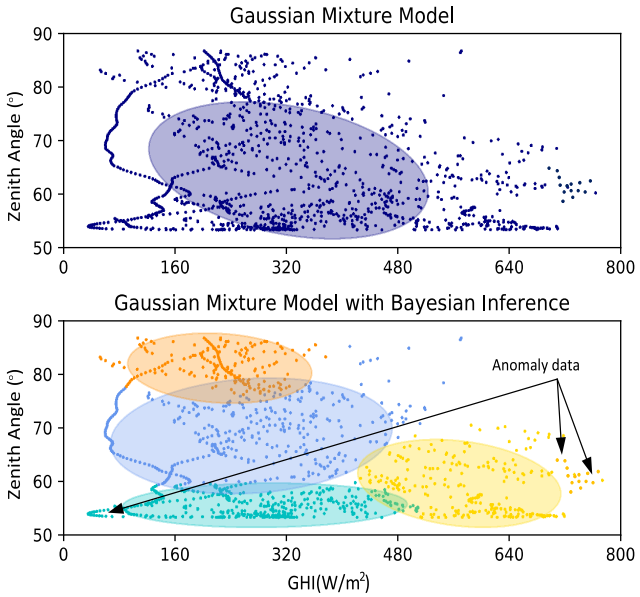


FIGURE 9. Anomaly detection for the GHI and Zenith Angle data.

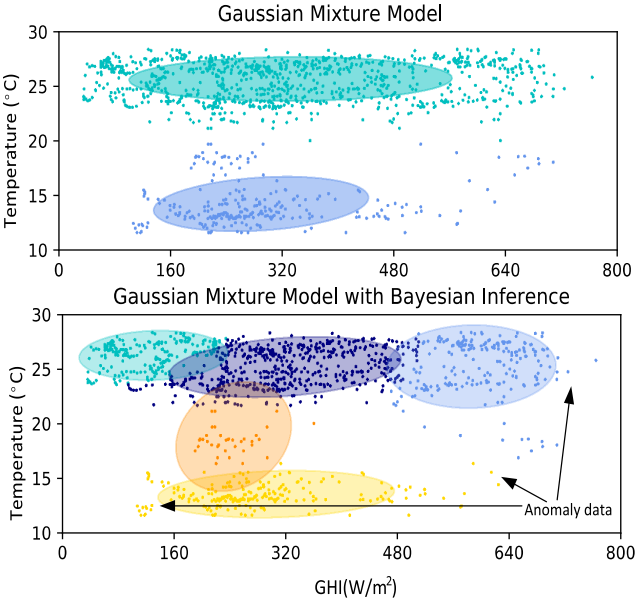


FIGURE 10. Anomaly detection for the GHI and Temperature data.

B. CNN BASED REGRESSION

Considering the limitation of the computer memory (64 GB), the data (including the sky images and corresponding GHI values) of 1, 2, 3, 4, and 5 days are used as training data, separately, and 70 randomly selected sky images and corresponding GHI values are used as testing data. As discussed above, the proposed CNN based image regression approach aims to build a mapping relationship from images to GHI values $h(\cdot)$: $\mathcal{X} \mapsto \mathcal{Y}$.

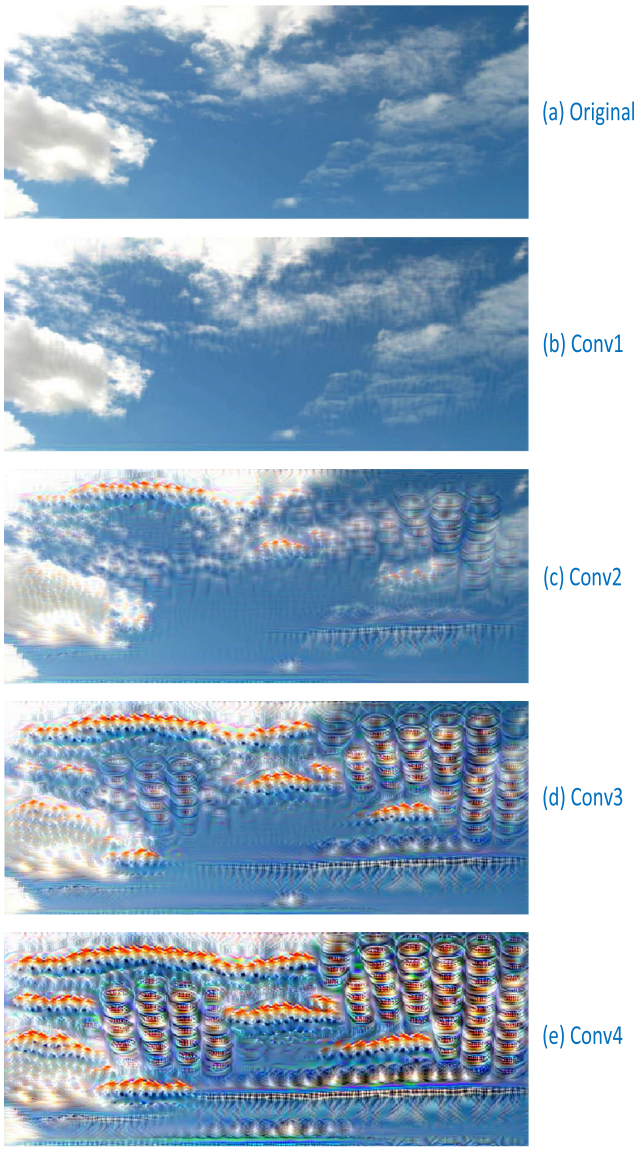


FIGURE 11. Feature expression of CNN (a) the original sky image demo, (b) the feature of conv1, (c) the feature of conv2, (d) the feature of conv3, (e) the feature of conv4.

1) FEATURE ANALYSIS

To illustrate clearly the operation manner of CNN, a demo of sky image is get from [36], and its feature analysis or content reconstruction is shown in Fig. 11. In Fig. 11(a), the original figure is shown, which is extracted from the beginning side of the flowchart in Fig. 7. Fig. 11(b) is almost identical with Fig. 11(a), and the small and tiny crinkles indicates the feature and edges of the clouds are extracted by the CNN. In Fig. 11(c), it is clearly that the edges of the clouds are selected by the crinkles, which means the features of the major parts are identified and located in the image. Furthermore, in Fig. 11(d) and Fig. 11(e), the features of the small parts are also identified and located. In brief, with deeper layers in the CNN, more and more features are extracted in differently layers gradually. With the extracted features, the regression analysis is introduced in next step.

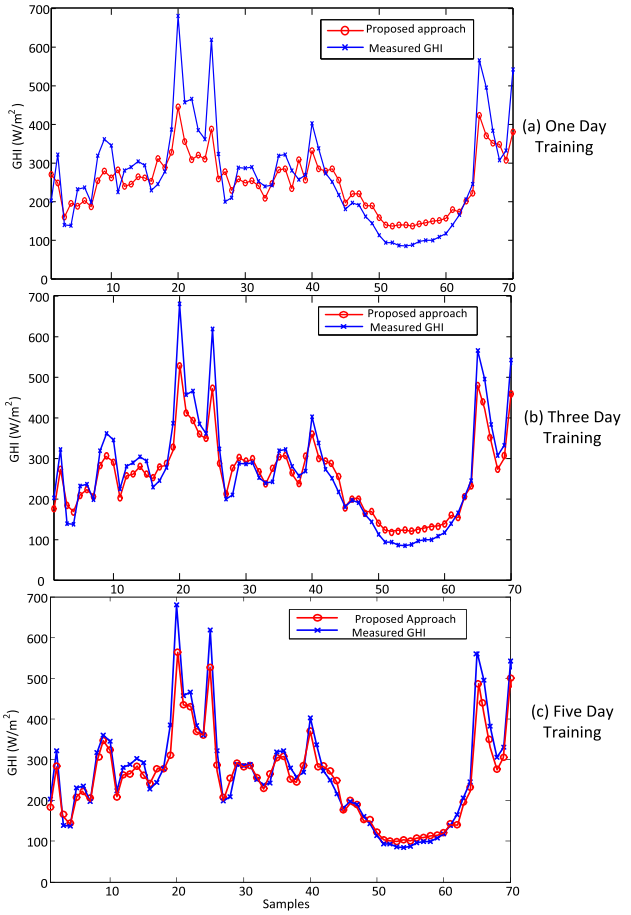


FIGURE 12. Image regression for GHI with different training data length: (a) 1 day data as training data, (b) 3 day data as training data, and (c) 5 day data as training data.

2) REGRESSION ANALYSIS

As shown in Fig.12, the top image regression results of training data with 1, 3, and 5 days are presented in red curves, and the testing data are presented with blue curves. The results of 1 day training data and testing data are presented in Fig. 12 (a), which contains a relatively big errors. Specifically, as shown with the blue curve, there are 4 peaks located in samples 20, 25, 65, and 70, which indicate high solar irradiance at these time slots. However, with 1 day training data, the results of the proposed approach are much lower, which indicates a big mismatch between the proposed approach and measured data. In addition, in the samples between 45 to 65, there is a big valley in the testing GHI values. However, as shown in Fig. 12 (a), the proposed approach also contains a big error from the measured GHI values. In Fig. 12 (a), the mean absolute percentage error (MAPE) is 21.8%, which indicates the 1 day training data is not sufficient to achieve a acceptable result.

Compared with Fig. 12 (a), the results presented in Fig. 12 (b) and Fig. 12 (c) with 3 and 5 days training data contain smaller errors. In Fig. 12 (b), it is obviously that the peak errors in samples 20, 25, 65, and 70 are much smaller than Fig. 12 (a), and the valley errors from samples

45 to 65 are also smaller than Fig. 12 (a). In Fig. 12 (c), the best regression results are presented with 5 days training data. The peak errors and valley errors are very small, and the blue curve and red curve are almost identical in the rest samples, which present are very accurate image regression results for solar irradiance capturing. In Fig. 12 (b) and Fig. 12 (c), the MAPE are 12.6% and 8.1%, which indicates the regression error is decreasing with the increasing training data length.

C. COMPARISON AND DISCUSSION

To comprehensive evaluate the proposed approach, several high impacted algorithms are investigated, and the comparison and analysis are shown as follows.

1) BENCHMARK COMPARISON

In [62], the persistence approach is introduced to evaluate the solar irradiance, which is a widely used approach to build the benchmark data [61]. In this paper, we also implement the persistence approach to estimate the solar irradiance. The MAPE of the persistence approach is 27.8% and root-mean-square error (RMSE) is 29.2%, which are much larger than the proposed approach with MAPE 8.1% and RMSE 8.7%. In [61], [65], the auto-regression integrated moving average (ARIMA) technique is also a widely used approach for solar irradiance forecasting, which requires stationary for the time series. For the cloudy days data, the MAPE and the RMSE of the ARIMA approach are 31.2% and and 32.7%, which are also very high.

2) OTHER APPROACHES COMPARISON

In [59], a image regression approach is proposed to predict solar irradiance with sky image, which consists of feature extracting and regression model with promising results. The MAPE 8.1% of the proposed approach is lower than 21.91% with approach in [59]. In [60], a statistic based artificial neural network approach is used to provide solar irradiance forecasting with minimum root-mean-square error (RMSE) about 15%, which is higher than our proposed approach with 8.7%. In the review paper [61], a lot of recent solar irradiation capturing and forecasting approach are collected and analyzed, and the best RMSE is about 10%, which is also higher than the proposed approach. Recently, in [63], a short-term solar irradiance forecasting approach is proposed with satellite and model coupling. In this approach, the sky images are captured by geostationary satellite and the typical errors ranges are from 8.5% to 17.2%, which is very close but still a little higher than the proposed approach. Furthermore, compared with the high-cost geostationary satellites, the proposed approach only requires the sky images, which can be easily captured by sky imagers, cameras, and even cellphones. In [64], several machine learning based approaches are used such as leapForward model and Spikeslab model to forecast the solar irradiance. Compared with the proposed approach, this proposed approach uses the measured solar irradiance as input, which indicates this method is heavily relies on the expensive solar irradiance measurement devices and not easy to extend on large-scale solar irradiance capturing. In [66], a solar irradiance forecasting approach is proposed with wavelet-based feature extraction, and this approach is focusing on the clear-sky days. Considering the large deviation profile of the PV output caused by clouds, the proposed approach is investigating in the cloudy days, which have larger impacts on power system operations. The randomness and stochastic characteristics of the cloud movements also increase the difficulties of the regression.

In this paper, we only investigating in the sky images taken with visible light as a primer exploration, which indicates other images such as 10-12 μm normalization radiance and R/B ratio mask are ignored. However, according to [59]–[61], these images contains useful information and can be treated as a good assistance information for image regression. Beside the Zenith angle and temperature, the other parameters such as humidity, air condition, and solar periods can also be seemed as assistance inputs for the proposed regression approach. In this primer exploration, we only collected three month data in the Autumn of 2016 and Summer of 2017 to validate the proposed approach. In the next step, we will collected at least three years data to further investigate the solar image regression.

As discussed above, the main advantages of the proposed approach is shown as following:

1) Compared with the traditional solar irradiation capturing approaches, the proposed approach doesn't require the expensive high-resolution image captured by geostationary satellites. The data collection cost (sky image) of the proposed approach is lower, which indicates the proposed approach is convenient to widely use and extend for very large-scale solar irradiance monitoring.
2) Compared with the results of benchmark and traditional machine learning approaches, the proposed approach provides an end-to-end manner to capture the solar irradiances. This not only reduces the information loss between the feature engineering and regression processing, but also improve the learning efficiency of the proposed approach.
3) In power system operations, the clouds movements causes a lot random deviations of PV outputs, which dramatically impact the health of the power systems. The proposed approach is focusing on the cloudy days solar irradiance capturing, which can be easily extended to PV output regression and power system stability forecasting in the next step research.

In addition, as a deep learning approach, CNN is widely used in a lot of areas such as medical, auto-driving, traffic, robotics, etc.. How to built a deep learning work automatically, how to reduce the training time or high efficient training the deep learning network, and how to reduce the dependency on big training data are the interesting topics for future research.

V. CONCLUSION

In this paper, a novel solar irradiance capturing approach is proposed with a CNN framework. For data cleaning, a Bayesian inference based automatic anomaly data detection is designed to eliminate the bad data in a nonparametric manner. Based on the existing CNN architecture, a proposed CNN is used to capture the GHI with the sky images. Compared with the traditional approaches, the proposed approach is accurate, flexible and convenient to be widely deployed for large-area solar irradiance capturing. Furthermore, the proposed approach can also be used as an automatic and multi-functional platform for other image regression projects such as tidal and geothermal power estimation.

Deep learning is an emerging technology with powerful ability in many aspects; however, it also contains many problems such as initial point setting, optimization, and architecture design. In the next step, we will collect more data, and focusing on improving the training efficiency, reducing the network complexity, and increasing its capability to provide more useful information for power system operation and control.

ACKNOWLEDGMENT

The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.

REFERENCES

[1] M. Cui, J. Zhang, A. Florita, B.-M. Hodge, D. Ke, and Y. Sun, "Solar power ramp events detection using an optimized swinging door algorithm," in *Proc. ASME 41st Design Autom. Conf.*, Aug. 2015, pp. 1–10.

[2] H. Jiang, Y. Li, Y. Zhang, J. J. Zhang, D. W. Gao, E. Muljadi, and Y. Gu, "Big data-based approach to detect, locate, and enhance the stability of an unplanned microgrid islanding," *J. Energy Eng.*, vol. 143, no. 5, Oct. 2017, Art. no. 04017045.

[3] X. Fang, Q. Hu, F. Li, B. Wang, and Y. Li, "Coupon-based demand response considering wind power uncertainty: A Strategic bidding model for load serving entities," *IEEE Trans. Power Syst.*, vol. 31, no. 2, pp. 1025–1037, Mar. 2016.

[4] Y. Luo, M. Andrus, J. Lian, S. K. Srivastava, and D. Cartes, "Power impact analysis of pulse power loads in the future integrated shipboard power system," in *Proc. ASNE Electr. Electr. Mach. Technol. Symp.*, Philadelphia, PA, USA, May 2012, pp. 1–5.

[5] X. Geng and L. Xie, "A data-driven approach to identifying system pattern regions in market operations," in *Proc. IEEE Power Energy Soc. General Meeting*, Jul. 2015, pp. 1–5.

[6] W. Yan, W. Gao, T. Gao, D. W. Gao, S. Yan, and J. Wang, "Distributed cooperative control of virtual synchronous generator based microgrid," in *Proc. IEEE Int. Conf. Electro Inf. Technol. (EIT)*, May 2017, pp. 506–511.

[7] Y. Gu, H. Jiang, Y. Zhang, J. J. Zhang, T. Gao, and E. Muljadi, "Knowledge discovery for smart grid operation, control, and situation awareness—A big data visualization platform," in *Proc. North Amer. Power Symp. (NAPS)*, Sep. 2016, pp. 1–6.

[8] T. P. Ackerman and G. M. Stokes, "The atmospheric radiation measurement program: To predict reliably what increased greenhouse gases will do to global climate, we have to understand the crucial role of clouds," *Phys. Today*, vol. 56, no. 1, pp. 38–44, 2003.

[9] T. L. Stoffel, I. Reda, D. R. Myers, D. Renne, S. Wilcox, and J. Treadwell, "Current issues in terrestrial solar radiation instrumentation for energy, climate, and space applications," *Metrologia*, vol. 37, no. 5, pp. 399–402, Oct. 2000.

[10] Y. Xie, M. Sengupta, and J. Dudhia, "A fast all-sky radiation model for solar applications (FARMS): Algorithm and performance evaluation," *Solar Energy*, vol. 135, pp. 435–445, Oct. 2016.

[11] G. Guo, Y. Fu, C. Dyer, and T. Huang, "Image-based human age estimation by manifold learning and locally adjusted robust regression," *IEEE Trans. Image Process.*, vol. 17, no. 7, pp. 1178–1188, Jul. 2008.

[12] M. S. Nixon and A. S. Aguado, *Feature Extraction & Image Processing for Computer Vision*. New York, NY, USA: Academic, 2012.

[13] H. Jiang, Y. Zhang, E. Muljadi, J. J. Zhang, and D. W. Gao, "A short-term and high-resolution distribution system load forecasting approach using support vector regression with hybrid parameters optimization," *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 3341–3350, Jul. 2018.

[14] Y. Gu, H. Jiang, J. J. Zhang, Y. Zhang, H. Wu, and E. Muljadi, "Multi-timescale three-phase unbalanced distribution system operation with variable renewable generations," *IEEE Trans. Smart Grid*, vol. 10, no. 4, pp. 4497–4507, Jul. 2019.

[15] M. Pipattanasomporn, H. Feroze, and S. Rahman, "Multi-agent systems in a distributed smart grid: Design and implementation," in *Proc. IEEE/PES Power Syst. Conf. Exposit.*, Mar. 2009, pp. 1–8.

[16] 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 Trans. Smart Grid*, vol. 5, no. 6, pp. 2947–2956, Nov. 2014.

[17] X. Geng and L. Xie, "Learning the LMP-load coupling from data: A support vector machine based approach," *IEEE Trans. Power Syst.*, vol. 32, no. 2, pp. 1127–1138, May 2017.

[18] Q. Lu and K.-D. Kim, "A mixed integer programming approach for autonomous and connected intersection crossing traffic control," in *Proc. IEEE 88th Veh. Technol. Conf. (VTC-Fall)*, Aug. 2018, pp. 1–6.

[19] H. Jiang, X. Dai, D. W. Gao, J. 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 Trans. Smart Grid*, vol. 7, no. 5, pp. 2525–2536, Sep. 2016.

[20] Q. Lu, "Safe and efficient intersection control of connected and autonomous intersection traffic," Jan. 2018, *arXiv:1801.09361*. [Online]. Available: https://arxiv.org/abs/1801.09361

[21] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[22] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[23] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, Mar. 2003.

[24] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. New York, NY, USA: Springer-Verlag, 2006.

[25] D. M. Blei and M. I. Jordan, "Variational inference for Dirichlet process mixtures," *Bayesian Anal.*, vol. 1, no. 1, pp. 121–143, Mar. 2006.

[26] K. Yoshii and M. Goto, "A nonparametric Bayesian multipitch analyzer based on infinite latent harmonic allocation," *IEEE Trans. Audio Speech Language Process.*, vol. 20, no. 3, pp. 717–730, Mar. 2012.

[27] Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei, "Sharing clusters among related groups: Hierarchical Dirichlet processes," in *Proc. Adv. Neural Inf. Process. Syst.*, 2005, pp. 1385–1392.

[28] 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 *Proc. 48th Asilomar Conf. Signals, Syst. Comput.*, Nov. 2014, pp. 530–534.

[29] Q. Lu and K.-D. Kim, "A genetic algorithm approach for expedited crossing of emergency vehicles in connected and autonomous intersection traffic," *J. Adv. Transp.*, vol. 2017, pp. 1–14, 2017.

[30] C. Feng, M. Cui, B.-M. Hodge, and J. Zhang, "A data-driven multi-model methodology with deep feature selection for short-term wind forecasting," *Appl. Energy*, vol. 190, pp. 1245–1257, Mar. 2017.

[31] M. Cui, J. Zhang, A. R. Florita, B.-M. Hodge, D. Ke, and Y. Sun, "An optimized swinging door algorithm for identifying wind ramping events," *IEEE Trans. Sustain. Energy*, vol. 7, no. 1, pp. 150–162, Jan. 2016.

[32] K. Zhang, Z. Yang, H. Liu, T. Zhang, and T. Başar, "Fully decentralized multi-agent reinforcement learning with networked agents," Feb. 2018, *arXiv:1802.08757*. [Online]. Available: https://arxiv.org/abs/1802.08757

[33] K. Zhang, A. Koppel, H. Zhu, and T. Başar, "Global convergence of policy gradient methods to (almost) locally optimal policies," Jun. 2019, *arXiv:1906.08383*. [Online]. Available: https://arxiv.org/abs/1906.08383

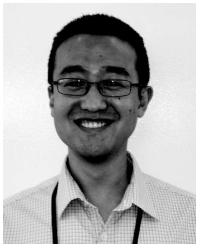
[34] Q. Lu and K.-D. Kim, "Autonomous and connected intersection crossing traffic management using discrete-time occupancies trajectory," *Appl. Intell.*, vol. 49, no. 5, pp. 1621–1635, May 2019.
[35] Q. Lu and K.-D. Kim, "Intelligent intersection management of autonomous traffic using discrete-time occupancies trajectory," *J. Traffic Logistics Eng.*, vol. 4, no. 1, pp. 1–6, 2016.
[36] *Wikipedia*. Accessed: May 2017. [Online]. Available: https://en.wikipedia.org/wiki/Sky
[37] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.
[38] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," Sep. 2014, *arXiv:1409.1556*, [Online]. Available: https://arxiv.org/abs/1409.1556
[39] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.
[40] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[41] D. Cao, Z. Lei, Z. Zhang, J. Feng, and S. Z. Li, "Human age estimation using ranking SVM," in *Proc. CCBR*. New York, NY, USA: Springer, 2012, pp. 324–331.
[42] Z. Niu, M. Zhou, L. Wang, X. Gao, and G. Hua, "Ordinal regression with multiple output CNN for age estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 4920–4928.
[43] Y. Zhang, R. Yang, K. Zhang, H. Jiang, and J. J. Zhang, "Consumption behavior analytics-aided energy forecasting and dispatch," *IEEE Intell. Syst.*, vol. 32, no. 4, pp. 59–63, Aug. 2017.
[44] H. Jiang and Y. Zhang, "Short-term distribution system state forecast based on optimal synchrophasor sensor placement and extreme learning machine," in *Proc. IEEE Power Energy Soc. General Meeting (PESGM)*, Jul. 2016, pp. 1–5.
[45] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," Feb. 2015, *arXiv:1502.03167*. [Online]. Available: https://arxiv.org/abs/1502.03167
[46] J. Zhao, G. Zhang, K. Das, G. N. Korres, N. M. Manousakis, A. K. Sinha, and Z. He, "Power system real-time monitoring by using PMU-based robust state estimation method," *IEEE Trans. Smart Grid*, vol. 7, no. 1, pp. 300–309, Jan. 2016.
[47] J. Zhao, M. Netto, and L. Mili, "A robust iterated extended Kalman filter for power system dynamic state estimation," *IEEE Trans. Power Syst.*, vol. 32, no. 4, pp. 3205–3216, Jul. 2017.
[48] F. Ding, H. Jiang, and J. Tan, "Automatic distribution network reconfiguration: An event-driven approach," in *Proc. IEEE Power Energy Soc. General Meeting (PESGM)*, Jul. 2016, pp. 1–5.
[49] D. Görür and C. Edward Rasmussen, "Dirichlet process Gaussian mixture models: Choice of the base distribution," *J. Comput. Sci. Technol.*, vol. 25, no. 4, pp. 653–664, Jul. 2010.
[50] B. A. Frigyik, A. Kapila, and M. R. Gupta. (2010). *Introduction to the Dirichlet Distribution and Related Processes*. [Online]. Available: http://mayagupta.org/publications/FrigyikKapilaGuptaIntroToDirichlet.pdf
[51] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[52] X. Zhou, M. Farivar, and L. Chen, "Pseudo-gradient based local voltage control in distribution networks," in *Proc. 53rd Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Sep. 2015, pp. 173–180.
[53] E. Skoplaki and J. Palyvos, "On the temperature dependence of photovoltaic module electrical performance: A review of efficiency/power correlations," *Solar Energy*, vol. 83, no. 5, pp. 614–624, May 2009.
[54] R. Marquez and C. F. Coimbra, "Intra-hour DNI forecasting based on cloud tracking image analysis," *Solar Energy*, vol. 91, pp. 327–336, May 2013.
[55] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, Jan. 2015.
[56] A. Dai, T. R. Karl, B. Sun, and K. E. Trenberth, "Recent trends in cloudiness over the united states: A tale of monitoring inadequacies," *Bull. Amer. Meteorolog. Soc.*, vol. 87, no. 5, pp. 597–606, May 2006.
[57] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.
[58] K. Simonyan, A. Vedaldi, and A. Zisserman, "Deep inside convolutional networks: Visualising image classification models and saliency maps," Dec. 2013, *arXiv:1312.6034*. [Online]. Available: https://arxiv.org/abs/1312.6034
[59] C.-L. Fu and H.-Y. Cheng, "Predicting solar irradiance with all-sky image features via regression," *Solar Energy*, vol. 97, pp. 537–550, Nov. 2013.
[60] R. Marquez and C. F. Coimbra, "Forecasting of global and direct solar irradiance using stochastic learning methods, ground experiments and the NWS database," *Solar Energy*, vol. 85, no. 5, pp. 746–756, May 2011.
[61] M. Diagne, M. David, P. Lauret, J. Boland, and N. Schmutz, "Review of solar irradiance forecasting methods and a proposition for small-scale insular grids," *Renew. Sustain. Energy Rev.*, vol. 27, pp. 65–76, Nov. 2013.
[62] K. Jan, *Solar Energy Forecasting and Resource Assessment*. New York, NY, USA: Academic, 2013.
[63] S. D. Miller, M. A. Rogers, J. M. Haynes, M. Sengupta, and A. K. Heidinger, "Short-term solar irradiance forecasting via satellite/model coupling," *Solar Energy*, vol. 168, pp. 102–117, Jul. 2018.
[64] A. Sharma and A. Kakkar, "Forecasting daily global solar irradiance generation using machine learning," *Renew. Sustain. Energy Rev.*, vol. 82, pp. 2254–2269, Feb. 2018.
[65] D. Yang, J. Kleissl, C. A. Gueymard, H. T. Pedro, and C. F. Coimbra, "History and trends in solar irradiance and PV power forecasting: A preliminary assessment and review using text mining," *Solar Energy*, vol. 168, pp. 60–101, Jul. 2018.
[66] T. Zhu, L. Xie, H. Wei, H. Wang, X. Zhao, and K. Zhang, "Clear-sky direct normal irradiance estimation based on adjustable inputs and error correction," *J. Renew. Sustain. Energy*, vol. 11, no. 5, Sep. 2019, Art. no. 056101.

HUAIGUANG JIANG (Member, IEEE) received the Ph.D. degree in electrical and computer engineering from the University of Denver, Denver, CO, USA, in 2015. He is currently with the National Renewable Energy Laboratory, Golden, CO. His research interests are in machine learning, data analysis, optimal control, renewable energy integration, smart grid, and transportation system analysis.

YI GU is currently pursuing the Ph.D. degree in electrical and computer engineering with the University of Denver, Denver, CO, USA. She is also an Intern with the National Renewable Energy Laboratory. Her research interests are in machine learning, data analysis, data visualization, computer vision, natural language processing, optimal control, and smart grid.

YU XIE worked with Texas A&M University and the Brookhaven National Laboratory. He has more than ten years of research experience. He is currently a Senior Scientist with the National Renewable Energy Laboratory (NREL), where he has been leading and participating in multiple research projects sponsored by the U.S. Department of Energy, NREL, and industry. He has authored or coauthored more than 20 journal publications. He specializes in solar resource assessment and forecasting. His specific topics of research concern developing radiative transfer models to accurately compute solar resource for all-sky conditions, predicting solar radiation and power for multiple timescales, and retrieving cloud macrophysical and microphysical properties using ground and satellite-based observations.

RUI YANG (Member, IEEE) received the B.E. degree in electrical engineering from Tsinghua University, Beijing, China, in 2009, and the Ph.D. degree in electrical and computer engineering from Carnegie Mellon University, Pittsburgh, PA, USA, in 2014. She is currently a Senior Research Engineer with the National Renewable Energy Laboratory, Power Systems Engineering Center, Golden, CO, USA. Her research interests include advanced data analytics, machine learning, and optimization for power systems applications.



YINGCHEN ZHANG (Senior Member, IEEE) received the B.S.E.E. degree from Tianjin University, Tianjin, China, in 2003, and the Ph.D. degree in electrical engineering from Virginia Polytechnic Institute and State University, Blacksburg, VA, USA, in 2010. He is currently with the National Renewable Energy Laboratory, Golden, CO, 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.