

# Solar Power Prediction Based on Satellite Images and Support Vector Machine

Han Seung Jang, *Student Member, IEEE*, Kuk Yeol Bae, *Student Member, IEEE*,  
Hong-Shik Park, *Member, IEEE*, and Dan Keun Sung, *Fellow, IEEE*

**Abstract**—Penetration of solar energy into main grid has gradually increased in recent years due to a growing number of large-scale photovoltaic (PV) farms. The power output of these PV farms may fluctuate due to a wide variability of meteorological conditions, and, thus, we need to compensate for this effect in advance. In this paper, we propose a solar power prediction model based on various satellite images and a support vector machine (SVM) learning scheme. The motion vectors of clouds are forecasted by utilizing satellite images of atmospheric motion vectors (AMVs). We analyze 4 years' historical satellite images and utilize them to configure a large number of input and output data sets for the SVM learning. We compare the performance of the proposed SVM-based model, the conventional time-series model, and an artificial neural network (ANN) model in terms of prediction accuracy.

**Index Terms**—Solar power, Irradiance, Prediction, Forecasting, Photovoltaic, Satellite images, Support vector machine, Machine learning.

## I. INTRODUCTION

THE world has faced a serious depletion problem of natural resources and a climate change problem due to an overuse of fossil fuels, and, thus, we need to take more alternative energy sources, so called 'renewable energy sources' instead of fossil fuels. Penetration of renewable energy sources into main grid has gradually increased in recent years and this penetration is expected to increase more rapidly until 2030. Focusing on the solar energy in South Korea, the penetration ratio of solar energy among all available energy sources is expected to gradually increase up to 14.1% until 2035 [1].

Large-scale photovoltaic (PV) farms have been widely deployed in the world [2]. Germany, China, and the U.S. have installed a number of large-scale PV farms and South Korea installed about 467 MW PV plants in 2013 [1]. Due to a growing number of large-scale PV farms, the percentage of solar power in total power grid keeps increasing and the power output of these PV farms may fluctuate due to a wide variability of meteorological conditions [3]. Accordingly, large-scale PV farm operators need to obtain more accurate prediction on how much energy can be produced from their PV farms so that they can trade solar power on the energy

market. In this situation, solar power prediction technologies play a very important role since depending on the accuracy of prediction, the amount of economic benefits from solar energy is significantly different.

There have been several studies on prediction technologies for solar irradiance or PV power generation [4], [5]. We can categorize these studies into three major categories. The first category includes solar irradiance or cloud index prediction technologies based on satellite images [6]–[8]. The variability of solar irradiance is mainly determined by the amount and motion of clouds, which can be detected using satellite images, and partly determined by the amount of aerosol [9], [10]. Meteorological satellite images are a useful source of detection of cloud-motion and the overall meteorological changes. Solar irradiance predictions using satellite images based on cloud motion vectors (CMVs) were proposed in [6], [11], [12]. The second category includes cloud detection technologies for solar irradiation prediction [13]–[16]. Image processing methodologies using total sky imagers (TSIs) to analyze cloud movement and generate short-term forecasts of solar irradiance at the ground level were proposed in [13], [15], [17]. The third category includes solar power prediction technologies based on machine learning schemes. Time series technologies such as ARMA and ARIMA were used for solar power prediction in [18], [19] and a spatial-temporal autoregressive with exogenous input (ARX) model was also proposed in [20]. However, these time series based prediction technologies have a limitation on non-linear parameters such as cloud movement and meteorological variables. Hence, artificial neural network (ANN) and support vector machine (SVM) schemes have been used in forecasting the global and horizon solar irradiance and power generation for PV systems [21]–[23].

There have been several studies on solar irradiance predictions based on the SVM scheme [22], [24]–[26]. These studies mainly utilized meteorological or climate data such as historical temperature and irradiance data in the SVM-based prediction models. However, in this paper, we propose a solar power prediction model based on various satellite images such as atmospheric motion vector (AMV), clouds, and irradiance images, and an SVM learning scheme. The advantage of utilizing satellite images on the solar power prediction is a macroscopic observation of the amount and movement of clouds. Furthermore, the SVM learning scheme can achieve high performance by minimizing the prediction error and maximizing the generalization ability of the prediction model. We first propose an extraction method of AMV from satellite images, and then forecast the motion of clouds

Manuscript received May 06, 2015; revised November 19 2015; revised January 14 2016; accepted February 23 2016. This work was supported in part by grant No. EEWS-2016-N11160018 from Climate Change Research Hub Project of the KAIST EEWS Research Center and in part by the KUSTAR-KAIST Institute, under the R&D program supervised by KAIST, Korea.

The authors are with the School of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea (e-mail: {jhans, kybae, park1507, dksung}@kaist.ac.kr).

Digital Object Identifier

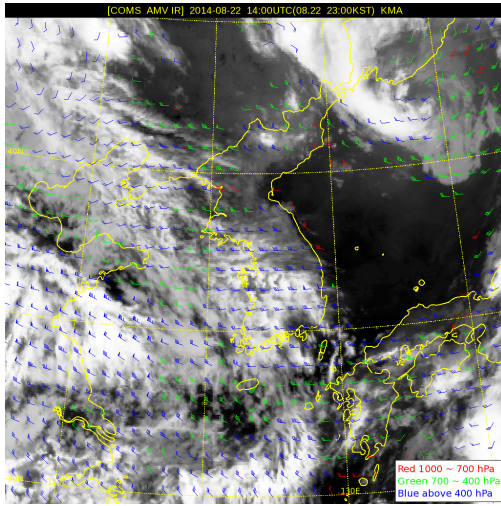


Fig. 1. An exemplary satellite image of AMV [27].

by utilizing these AMVs. We thoroughly analyze 4 years' historical satellite images and utilize them to configure a large number of input and output data sets for the SVM learning. As a result, the proposed SVM-based prediction model can simultaneously predict the future amount of clouds and solar irradiance on multiple sites of South Korea in the range of 15 to 300 minutes (intra-day). We show the prediction accuracy of the proposed model, compared with that of the conventional time-series and the ANN models.

The rest of this paper is organized as follows. In Section II, we introduce satellite images utilized in this paper. In Section III, we propose an AMV extraction method. We present the extraction method of cloud factors from satellite images in Section IV. In Section V, we address the training method with SVM regression. We evaluate the performance of the proposed SVM-based prediction model in Section VI. Finally, we draw conclusive remarks in Section VII.

## II. SATELLITE IMAGES

Meteorological satellite images for North-East Asia are provided by the Korea National Meteorological Satellite Center (NMSC) [27]. The NMSC launched a communication, ocean and meteorological satellite (COMS) on June 27, 2010, which is the Korea's first geostationary multi-purpose satellite and has performed meteorological and ocean observations and communications services. The major missions of the COMS include continuous monitoring of imagery and extraction of meteorological products, early detection of severe weather phenomena, and monitoring of climate change and atmospheric environment. The NMSC provides raw images, basic images, and processed images for North-East Asia and the Korean Peninsula. Each of satellite images for the Korean Peninsula has  $1024 \times 1024$  pixels and the ground resolution of each pixel is 1720.8 m. Most of images are updated every 15 minutes, but approximately 1 hour is delayed due to the procedure of communications and image processing. In this section, we introduce main satellite images used for this paper such as atmospheric motion vector (AMV) image, cloud analysis image, and irradiance image.

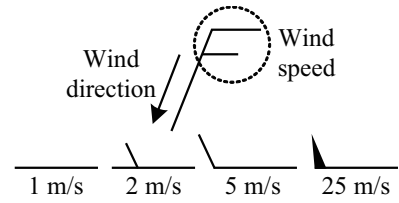


Fig. 2. A standard station model for wind direction and speed.

### A. Atmospheric Motion Vector (AMV) images

Atmospheric motion vector (AMV) provides the overall information of atmospheric motion and wind direction and speed of lower, middle, and upper wind fields [28]. Fig. 1 shows an exemplary satellite image of AMV. The red, green, and blue colors indicate the lower, middle, and upper wind fields, respectively. Fig. 2 shows a standard station model for the wind direction and speed. There exist various AMVs in three colors on a satellite image of Korean Peninsula and, thus, we should accurately extract those AMVs in a novel way of image processing. We will describe the AMV extraction method later in detail.

### B. Cloud analysis images

The NMSC especially provides a variety of cloud analysis images, which contain the amount of clouds, cloud shape, and cloud thickness images. The image of the amount of clouds shows the amount of clouds in the range of 0 ~ 100 on each pixel. The cloud shape image shows the shape of cloud based on the difference in the absorption of cloud particles and cloud temperature in the range of lower cloud (1 ~ 3), middle cloud (4), upper cloud (5 ~ 9), and mixed cloud (10) on each pixel. In the last, the cloud thickness image shows the thickness of clouds in the range of 0 ~ 100 on each pixel. These cloud analysis images are very important factors for the prediction of future condition of clouds on the target area, and are used for input variables of machine learning.

### C. Irradiance images

Irradiance images are also provided by the NMSC. These images are based on the amount of light intensity reflected by the ground and the irradiance is measured in Watt per square meter ( $W/m^2$ ) in the range of 0 ~ 1000  $W/m^2$ . The irradiance can be utilized together with a module temperature to forecast the photovoltaic (PV) power generation [29], [30].

## III. PROPOSED AMV EXTRACTION METHOD

In this section, we propose an AMV extraction method to gather the information of wind direction and speed. Before we address the proposed method, we define some variables and make notations. Fig. 3 shows a single AMV vector. The head, body, and tail are shown at the top of the AMV wing (red circle), at the bottom of the AMV wing (green circle), and at the end of the AMV line (blue circle), respectively. In addition, the middle points between the head and body, and between the head and tail are marked with orange and purple diamonds, respectively.

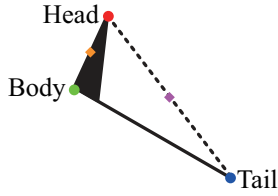


Fig. 3. A single AMV vector.

#### A. Procedure of AMV Extraction

The proposed AMV extraction method consists of four steps. The first step is to separate red, green, and blue pixels from an AMV satellite image and save their pixel coordinates into each repository of red AMVs, green AMVs, and blue AMVs. For better understanding, we explain the remaining procedure with respect to the blue AMVs.

The second step is to extract each single blue AMV from the repository of blue AMVs. At the beginning, we randomly select one pixel from the repository of blue AMVs and then calculate the distances between the selected pixel and the other pixels in the repository of blue AMVs. After the completion of distance calculations, we store the neighboring pixel with the minimum distance into a sub-repository of a single blue AMV and also assign this pixel as the next selected point. Every time when we store a new pixel into the sub-repository of the single blue AMV, the center coordinate of pixels in the sub-repository of the single blue AMV is calculated. We continue the same procedure by considering the following two constraints: 1) the distance between the pre-selected pixel and the next selected pixel should be less than a threshold and 2) the next selected pixel is close under a threshold to the center coordinate. If there are no other pixels satisfying the two constraints on the repository of blue AMVs, the extraction of the single blue AMV is finished, and we complete the sub-repository of the single blue AMV. Continuing this step, all sub-repositories of single blue AMVs are obtained from the repository of blue AMVs.

The third step is to find the head, body, and tail pixels from the sub-repositories of the single blue AMVs. In this step, we pick up the farthest two pixels from a sub-repository for assigning the head and tail and calculate their middle point (a purple diamond in Fig. 3). Except the two picked pixels, the farthest pixel from the middle point (purple diamond) in the sub-repository is assigned as the body of the single blue AMV. Then, throughout the distance comparison between the body and the two picked pixels, the shorter-distance pixel is assigned as the head and the longer-distance pixel is assigned as the tail. At the last of the third step, a wind direction vector  $\mathbf{d} = (d_x, d_y)$  is calculated from the body to tail, and the angle of the wind direction is calculated as  $\theta$ . We continue this procedure for the remaining sub-repositories.

The fourth step is to calculate the wind speed for each single blue AMV. The total number of pixels in a sub-repository of a single blue AMV is counted and then the middle point (an orange diamond in Fig. 3) between the head and body pixels is calculated. After that, we calculate the distances between the middle point (orange diamond) and each of pixels in the

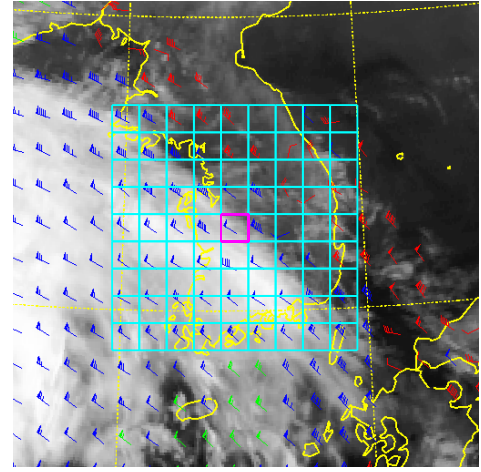


Fig. 4. The target and search areas on an AMV satellite image.

sub-repository of the single blue AMV. If the number of pixels whose distances from the middle point (orange diamond) are under a distance threshold is larger than 4 pixels, the wind speed is counted starting from 25 m/s, otherwise 5 m/s. The wind speed is finally calculated by additionally considering remaining speed bars (long bar: 5 m/s, short bar: 2 m/s).

#### IV. EXTRACTION OF CLOUD IMPACT FACTORS FROM SEARCH AREAS

First of all, we specify a target area and a search area. The target area is a place where we want to predict solar power. It has a target coordinate  $(x_{\text{target}}, y_{\text{target}})$ , width  $w_{\text{target}}$ , and height  $h_{\text{target}}$ . The search area is a neighboring zone of the target area and the  $i$ -th search area has a search coordinate  $(x_{\text{search},i}, y_{\text{search},i})$ , width  $w_{\text{search}}$ , and height  $h_{\text{search}}$ . Depending on the size of the target area, the accuracy and locality of the prediction may be different.

In this section, we address how to extract cloud factors from the search areas and characterize the cloud impact factors of each search area on the target area. Since we have three colors of AMV,  $k = R, G, \text{ or } B$ , according to the color notation and the index of search area, the wind direction, speed, and the angle of the single AMVs with color  $k$  on the  $i$ -th search area are denoted by  $\mathbf{d}_{i,k} = \{d_x(i, k), d_y(i, k)\}$ ,  $V_{i,k}$  (m/s), and  $\theta_{i,k}$ , respectively.

##### A. Calculation of cloud impact factor

We assume that the movement of clouds approximately follows the wind direction and speed. However, since we can conjecture that the speed of clouds and wind are different, we may find the optimal ratio of the cloud speed to the wind speed.  $\alpha^*$  denotes the optimal ratio. Another assumption is that we do not consider a new formation and disappearance of clouds. Fig. 4 shows the overall AMV image of the target area (the magenta box) and the search areas (the cyan boxes). On the  $i$ -th search area with color  $k$ , a set of average AMV parameters is calculated by a number of iterations and it is

denoted by  $\Omega_{i,k} = \{\bar{\mathbf{d}}_{i,k} = \{\bar{d}_x(i,k), \bar{d}_y(i,k)\}, \bar{V}_{i,k}, \bar{\theta}_{i,k}\}$ . With the set of parameters  $\Omega_{i,k}$ , we can expect the movement of the  $i$ -th search area after  $t$  minutes. The center coordinate of the  $i$ -th search area is updated as follows.

$$\begin{aligned} x_{i,k}(t) &= x_{\text{search},i} + \left\lceil \frac{\alpha^* \times \bar{V}_{i,k} \times 60 \times t \times \cos \bar{\theta}_{i,k}}{\varepsilon} \right\rceil, \\ y_{i,k}(t) &= y_{\text{search},i} + \left\lceil \frac{\alpha^* \times \bar{V}_{i,k} \times 60 \times t \times \sin \bar{\theta}_{i,k}}{\varepsilon} \right\rceil, \end{aligned}$$

where  $(x_{i,k}(t), y_{i,k}(t))$  and  $\varepsilon$  (meters) are the newly moved center coordinate of the  $i$ -th search area with color  $k$  after  $t$  minutes and the ground resolution of one pixel, respectively. In addition, if  $w_{\text{target}} = w_{\text{search}}$  and  $h_{\text{target}} = h_{\text{search}}$ , the amount of sharing between the target area and the moved search area is calculated as,

$$\zeta_{i,k}(t) = \frac{\max\{(w_{\text{target}} - A), 0\} \times \max\{(h_{\text{target}} - B), 0\}}{w_{\text{target}} \times h_{\text{target}}},$$

where  $A = |x_{i,k}(t) - x_{\text{target}}|$  and  $B = |y_{i,k}(t) - y_{\text{target}}|$ .

Cloud factors such as the amount of clouds and the shape of clouds are calculated in both of the target area and the search areas.  $C_i$  and  $S_i$  denote the average amount and shape of clouds on the  $i$ -th area, respectively. The three kinds of wind fields, R, G, and B, and the cloud shapes (1 ~ 10) should be matched in terms of altitudes. The matching rule is  $R = \{1, 2\}$ ,  $G = \{3, 4\}$ , and  $B = \{5, \dots, 10\}$ , and according to the matching rule, only one of the amount of sharing is activated, i.e., if  $S_i = 1$ ,  $\zeta_{i,R}(t)$  is activated, but  $\zeta_{i,G}(t) = \zeta_{i,B}(t) = 0$ .

Lastly, we can calculate the cloud impact factor of the  $i$ -th search area on the target area after  $t$  minutes as follows:

$$\gamma_i(t) = C_i \times \sum_{k=R,G,B} \zeta_{i,k}(t) \quad (1)$$

The total cloud impact factor on the target area after  $t$  minutes is calculated as

$$\Gamma(t) = \sum_{i=1}^I \gamma_i(t), \quad (2)$$

where  $I$  is the total number of search areas.

## V. TRAINING WITH SUPPORT VECTOR MACHINE REGRESSION

We can train our solar power prediction model based on the support vector machine (SVM) regression scheme with a large amount of historical satellite image data. Before we address the training and test procedure, we define some parameters. We define a given set of  $N$  input and output data pairs as  $\mathcal{D} = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_N, y_N)\}$ , and  $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$  and  $\mathcal{Y} = \{y_1, y_2, \dots, y_N\}$  denote the sets of input and output data, respectively. Here in  $\mathcal{X}$ ,  $\mathbf{x}_i = [x_{i,1}, x_{i,2}, \dots, x_{i,K}]^T$  is the  $i$ -th input column vector with total  $K$  variables. We use  $\mathcal{D}_{\text{training}} = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_M, y_M)\}$  which denotes the total  $M$  input and output data for training. The remaining  $(N - M)$  input and output data are used for test, and the set of test data is denoted by  $\mathcal{D}_{\text{test}} = \{(\mathbf{x}_{M+1}, y_{M+1}), (\mathbf{x}_{M+2}, y_{M+2}), \dots, (\mathbf{x}_N, y_N)\}$ .

The input and output data may have a complex nonlinear relationship with each other. This complex nonlinear relationship can be formulated by  $y_i = g(\mathbf{x}_i)$  and, thus, our main goal of machine learning (ML) is to find the most similar function  $f(\mathbf{x}_i) \approx g(\mathbf{x}_i)$ . In this paper, we will find a function  $f(\mathbf{x}_i)$  with the SVM regression.

### A. Support Vector Machine Regression

Support vector machine (SVM) is one of high-performance ML schemes, which was originally proposed by Vapnik [31]. SVM is designed to not only minimize the error, but also maximize the separation margin among different classes. SVM can be applied to regression methods in supervised learning [32]. In the SVM regression model, a regression function is formulated as  $f(\mathbf{x}_i) = \mathbf{w}^T \cdot \mathbf{x}_i + b$  with  $\mathbf{x}_i, \mathbf{w} \in \mathbb{R}^K$  and  $b \in \mathbb{R}$ , where  $\mathbf{w}^T \cdot \mathbf{x}_i$  is the inner product between  $\mathbf{w}$  and  $\mathbf{x}_i$ . Finding the optimal  $\mathbf{w}$  is the main goal of the SVM regression. The optimization problem for finding the optimal  $\mathbf{w}$  is formulated as [31]

$$\begin{aligned} \underset{\mathbf{w}}{\text{minimize}} \quad & \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^M (\xi_i + \xi_i^*) \\ \text{subject to} \quad & y_i - \mathbf{w}^T \cdot \mathbf{x}_i - b \leq \varepsilon + \xi_i, \forall i, \\ & \mathbf{w}^T \cdot \mathbf{x}_i + b - y_i \leq \varepsilon + \xi_i^*, \forall i, \\ & \xi_i, \xi_i^* \geq 0, \forall i, \end{aligned}$$

where  $(\xi_i, \xi_i^*)$ ,  $\varepsilon$ , and  $C$  are the slack variables, a precision parameter, and the constant determining the trade-off between the flatness of  $f$  and tolerance of  $\varepsilon$ . With a Lagrange function and a dual optimization problem, we obtain the optimal coefficient vector  $\mathbf{w}^* = \sum_{i=1}^M (\beta_i - \beta_i^*) \mathbf{x}_i$ , where  $\beta_i$  and  $\beta_i^*$  are the Lagrange multipliers. For a non-linear SVM regression model, we introduce a kernel method. Kernel  $\Phi(\cdot)$  expands the input data vector  $\mathbf{x}_i$  with  $K$  dimensions into the input data vector with  $K + D$  dimensions, i.e.,  $\Phi(\mathbf{x}_i) \in \mathbb{R}^{K+D}$ . Hence, the non-linear SVM regression function is formulated as  $f(\mathbf{x}_i) = \mathbf{w}^T \cdot \Phi(\mathbf{x}_i) + b$  with  $\Phi(\mathbf{x}_i), \mathbf{w} \in \mathbb{R}^{K+D}$  and  $b \in \mathbb{R}$ , and the optimal coefficient vector is found as  $\mathbf{w}^* = \sum_{i=1}^M (\beta_i - \beta_i^*) \Phi(\mathbf{x}_i)$ . In the last, the optimal non-linear SVM regression function is formulated as

$$\begin{aligned} f(\mathbf{x}) &= \sum_{i=1}^M (\beta_i - \beta_i^*) \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}) + b \\ &= \sum_{i=1}^M (\beta_i - \beta_i^*) \mathcal{K}(\mathbf{x}_i, \mathbf{x}) + b, \end{aligned} \quad (3)$$

where the radial basis function (RBF) kernel,  $\mathcal{K}(\mathbf{x}_i, \mathbf{x}) = \exp\{-\|\mathbf{x}_i - \mathbf{x}\|^2 / 2\sigma^2\}$ , is used for this paper. A method of finding  $b$  can be referred to [32].

### B. Training Method

We have two objectives of prediction. The first one is to predict the future amount of clouds at the target area. The other one is to predict the future irradiance at the target area. Depending on the prediction objectives, we need to configure the input and output data sets in a different way for ML

trainings. In this subsection, we address the training methods for each prediction model.

In the prediction of the amount of clouds, we configure the input data vector  $\mathbf{x}_C$  with a set of the current amount of clouds at the search areas  $\mathcal{C}$ , and the total cloud impact factor  $\Gamma(t)$  for the prediction horizon of  $t$  minutes. Therefore, the input column vector is configured as  $\mathbf{x}_C = [\mathcal{C}, \Gamma(t)]^T$  and the dimension of the vector is  $I+1$ , where  $I$  is the total number of the search areas. The amount of clouds at the target area after  $t$  minutes is set to a single output value, i.e.,  $y_C = C_{\text{target}}(t)$ . The amount of clouds can be used to determine the clear-sky index  $\eta$ , defined as the ratio of the irradiance  $\Psi_{\text{target}}(t)$  at the target area to the clear-sky irradiance  $\Psi_{\text{clear}}(t)$ . The amount of clouds  $C_{\text{target}}(t)$  has an approximately linear relationship with the clear-sky index  $\eta$  as follows [33]:

$$\eta = \frac{\Psi_{\text{target}}(t)}{\Psi_{\text{clear}}(t)} \approx 1 - \frac{C_{\text{target}}(t)}{100} \quad (4)$$

Thus, the irradiance of the target area after  $t$  minutes can be predicted as  $\hat{\Psi}_{\text{target}}(t) = \Psi_{\text{clear}}(t) \times (1 - \hat{C}_{\text{target}}(t)/100)$ .

In the irradiance prediction, we additionally insert the solar altitude angle  $\omega$ , which is calculated at the specific date and time, and the current irradiance of the target area  $\Psi_{\text{target}}(0)$  into the input data vector for the prediction of amount of clouds  $\mathbf{x}_C$  explained above since the solar altitude angle and the current irradiance are reference data for the prediction of the irradiance value. Hence, the input column vector for the prediction of irradiance is configured as  $\mathbf{x}_\Psi = [\mathcal{C}, \Gamma(t), \omega, \Psi_{\text{target}}(0)]^T$ , whose dimension is  $I+3$ , and the irradiance of the target area after  $t$  minutes is set to a single output value, i.e.,  $y_\Psi = \Psi_{\text{target}}(t)$ .

## VI. PERFORMANCE EVALUATION

In this section, we first investigate the statistics of the amount of clouds and irradiance in South Korea in order to understand the characteristics of both meteorological data. Then, we evaluate the performance of the proposed SVM-based prediction models for the amount of clouds and irradiance in terms of prediction accuracy metrics such as root mean square error (RMSE), mean relative error (MRE), and the coefficient of determination,  $R^2$ , which are defined, respectively, as follows:

- $\text{RMSE} = \sqrt{\frac{\sum_{i=M+1}^N (y_i - f(\mathbf{x}_i))^2}{N - M}},$
- $\text{MRE} [\%] = \frac{1}{N - M} \sum_{i=M+1}^N \frac{|y_i - f(\mathbf{x}_i)|}{\bar{y}} \times 100,$
- $R^2 = 1 - \frac{\sum_{i=M+1}^N (y_i - f(\mathbf{x}_i))^2}{\sum_{i=M+1}^N (y_i - \bar{y})^2},$

where  $\bar{y} = (\sum_{i=M+1}^N y_i)/(N - M)$ .

The parameters and values for the experiments are summarized in Table I. We utilize 4 years' (2011.04 ~ 2015.09) satellite images from the NMSC in order to configure a large number of data sets for the ML training and test. 70% and 30% of the data are utilized for the training and the test procedures, respectively. The center of the target area is

TABLE I  
PARAMETERS AND VALUES FOR THE EXPERIMENTS

Parameters	Values
The size of satellite image	1024 × 1024 pixels
The ground resolution of one pixel	1720.8 m
$w_{\text{search}} = h_{\text{search}}$ pixels	1, 3, 5, 11, 21, 41
Prediction horizon, $t$	15 ~ 300 minutes

set to the location of Korea Advanced Institute of Science and Technology (KAIST) with a coordinate of (539,581), Daejeon, South Korea. In this paper, the MATLAB-based LIBSVM (LIBrary for Support Vector Machines) [34] is used for developing the prediction models.

### A. Statistics of the amount of clouds and irradiance

We observe both the amount of clouds and irradiance [35] for 9 hours from 9:00 to 17:00 at the target area. Fig. 5 shows the monthly statistics of the amount of cloud and irradiance, respectively. We can classify the characteristics of the cloud amount of each month into three classes such as small, medium, and large amount of clouds based on the average amount of clouds of each month. January, March, and October belong to the small-amount class of clouds. February, April, May, September, November, and December belong to the medium-amount class of clouds. June, July, and August belong to the large-amount class of clouds, which shows a very large amount and high variation of clouds in summer seasons. In the statistics of irradiance, May and December show the highest and lowest average values at the target area, respectively.

Fig. 6 shows the hourly statistics of the amount of clouds and irradiance at the target area, respectively. The hourly statistics of the amount of clouds shows similar features over 9 hours. The hourly statistics of irradiance shows a general hourly irradiance pattern, which around 12 and 13 o' clocks have peak irradiance values.

### B. Prediction for future amount of cloud

Before making a prediction model for the amount of clouds, we need to find the optimal ratio  $\alpha^*$  of the cloud speed to wind speed, which is used to compute the cloud impact factors. To do this, we utilize the correlation values between the variation of the amount of clouds after  $t$  minutes and the cloud impact factors, i.e.,  $\rho(\alpha) = \text{corr}[\{C_{\text{target}}(t) - C_{\text{target}}(0)\}, \Gamma(t, \alpha)]$ . The correlation value  $\rho(\alpha)$  increases as the variation of the amount of clouds and the cloud impact factor show a similar increasing or decreasing pattern. Fig. 7 shows that the optimal ratio  $\alpha^*$  is found at the maximum correlation value.

Table II summarizes the prediction results for the amount of clouds with a prediction horizon of 60 minutes. The width and height of the target and search areas are the same as 41 pixels, each on this result. The proposed SVM-based prediction model for the amount of clouds outperforms the persistence model, a non-linear auto regressive (NAR) model, and an artificial neural network (ANN) model. The proposed prediction model yields the RMSE, MRE, and  $R^2$  values of 10.8661, 9.9677, and 0.9104, respectively. Especially, the persistence-based



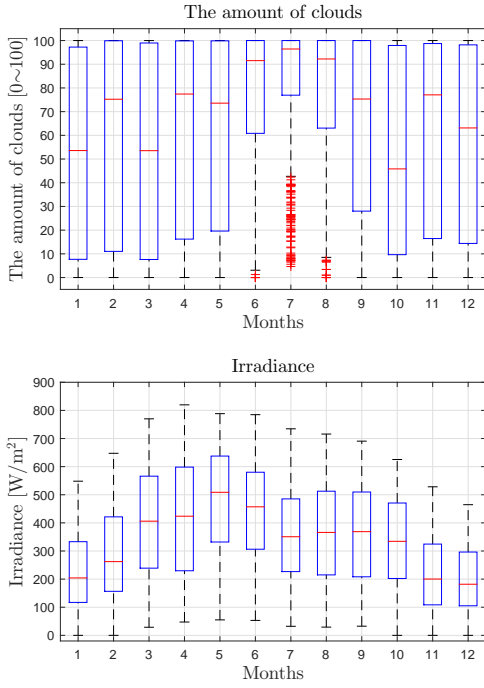


Fig. 5. Monthly statistics for the amount of clouds and irradiance

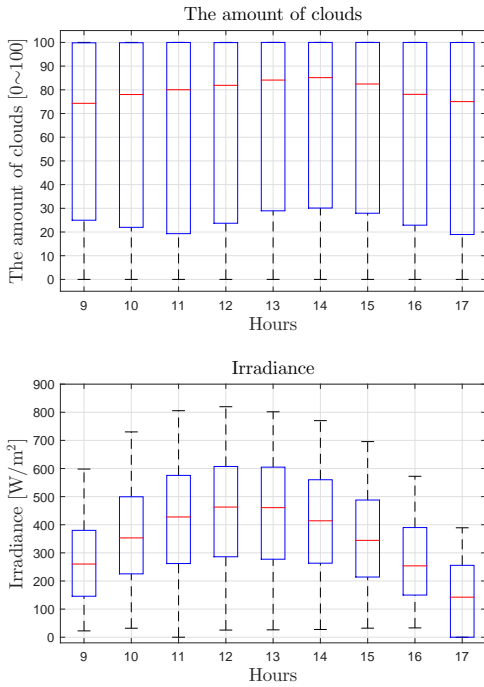


Fig. 6. Hourly statistics for the amount of clouds and irradiance

prediction model yields the lowest prediction accuracy for all metrics.

Fig. 8 compares the predicted amount of clouds of the proposed SVM-based model with the measured values. Even though some prediction errors occur, the prediction results of the proposed model generally agree with the measured values.

Fig. 9 shows the scatter plot of the measured and predicted amount of clouds for the proposed SVM-based prediction

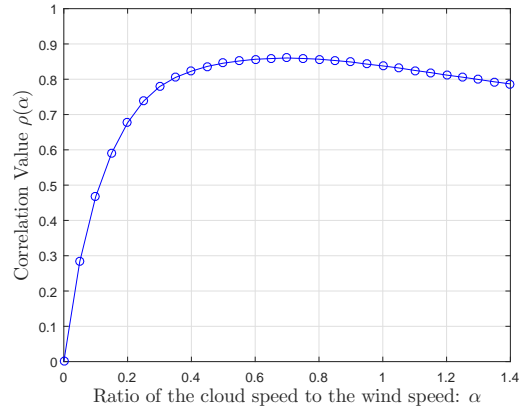


Fig. 7. Finding the optimal ratio  $\alpha^*$  of the cloud speed to the wind speed

TABLE II  
PREDICTION RESULTS FOR THE AMOUNT OF CLOUDS WITH  $t = 60$  MINUTES AND  $w_{\text{target}} = h_{\text{target}} = 41$  PIXELS

Schemes \ Metrics	RMSE	MRE (%)	$R^2$ (0~1)
Persistence	13.8489	12.7296	0.8554
NAR	12.2343	12.3699	0.8641
ANN	12.0336	12.6681	0.8774
SVM	10.8661	9.9677	0.9104

model. The more the scatter plot converges to the red line, the better the prediction model performs. It indicates that a larger amount of clouds (near 100) and a smaller amount of clouds (near 0) are very well predicted, while it is rather hard to predict the medium amount of clouds.

The distribution of prediction errors approximately follows the non-standardized Student's  $t$ -distribution [36], whose probability density function (PDF) is

$$p(x|\nu, \mu, \sigma) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2}) \sqrt{\nu\pi\sigma^2}} \left\{ 1 + \frac{1}{\nu} \left( \frac{x - \mu}{\sigma} \right)^2 \right\}^{-\frac{\nu+1}{2}},$$

where  $\Gamma(\cdot)$ ,  $\mu$ ,  $\sigma$ , and  $\nu$  denote the gamma function, the location parameter, the scale parameter, and the shape parameter, respectively. The mean and variance of the distribution are  $\mu$  and  $\sigma^2\nu/(\nu-2)$ . The  $\mu$ ,  $\sigma$ , and  $\nu$  values of prediction errors for the amount of clouds with a prediction horizon of 60 minutes are 0.2842, 3.6460, and 1.25532, respectively.

Table III summarizes the prediction results of the amount of clouds based on the proposed SVM-based model with various prediction time horizons. It is well known that the suitable forecast time horizon based on satellite images is up to 5 hours [4]. The very-short-term (15 and 30 minutes horizon) and short-term (60 minutes horizon) prediction for the amount of clouds show high prediction accuracy with  $R^2 = 0.9745$ , 0.9444, and 0.9104, while the prediction accuracy of 90 ~ 300 minutes horizons degrades to  $R^2 = 0.8261 \sim 0.5757$ , which comes from the fact that new formation and disappearance of clouds may occur and they are not considered in the proposed prediction model.

Another issue is locality in the prediction. Table IV summarizes the prediction results of the amount of clouds based

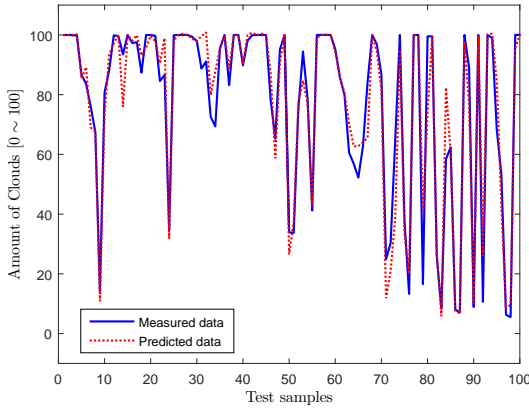


Fig. 8. The measured and predicted amount of clouds based on the proposed SVM-based model with  $t = 60$  minutes

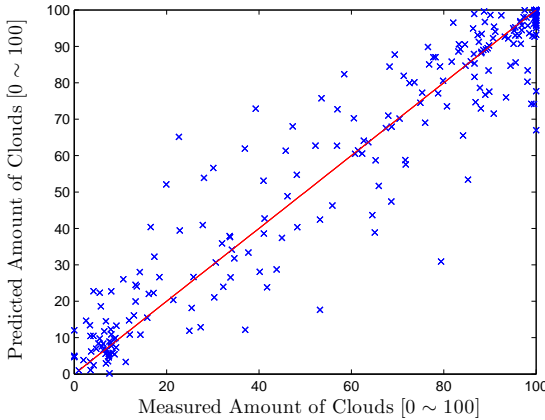


Fig. 9. Scatter plot of the measured and predicted amount of clouds based on the proposed SVM-based model with  $t = 60$  minutes

TABLE III  
PREDICTION RESULTS FOR THE AMOUNT OF CLOUDS FOR VARIOUS PREDICTION TIME HORIZONS WITH  $w_{\text{target}} = h_{\text{target}} = 41$  PIXELS

Horizons \ Metrics	RMSE	MRE (%)	$R^2$ (0~1)
15 minutes	5.7367	5.2825	0.9745
30 minutes	8.4336	7.1020	0.9444
60 minutes	10.8661	9.9677	0.9104
90 minutes	14.9205	13.4699	0.8261
120 minutes	16.8566	17.1327	0.7977
150 minutes	17.5458	17.8165	0.7646
180 minutes	19.5643	18.1852	0.7021
240 minutes	23.7025	25.0641	0.6049
300 minutes	24.7855	26.6654	0.5757

TABLE IV  
PREDICTION RESULTS FOR THE AMOUNT OF CLOUDS WITH VARIOUS RADII OF THE TARGET AREA AND  $t = 60$  MINUTES

Radius \ Metrics	RMSE	MRE (%)	$R^2$ (0~1)
35 km	10.8661	9.9677	0.9104
18 km	12.3090	10.8526	0.8880
9.4 km	13.4196	12.1586	0.8728
4.3 km	14.4074	13.1201	0.8487
2.6 km	15.0402	13.7125	0.8435
0.9 km	17.4361	15.1909	0.7952

TABLE V  
PREDICTION RESULTS FOR THE IRRADIANCE WITH  $t = 60$  MINUTES AND  $w_{\text{target}} = h_{\text{target}} = 41$  PIXELS

Schemes \ Metrics	RMSE ( $\text{W}/\text{m}^2$ )	MRE (%)	$R^2$ (0~1)
Persistence	57.5704	11.0767	0.8890
NAR	74.8116	15.1280	0.8172
ANN	65.5209	12.1202	0.8564
SVM	44.1390	7.7329	0.9420

on the proposed SVM-based model for various radii of the target area and  $t = 60$  minutes. The radii of the target area, 35 km, 18 km, 9.4 km, 4.3 km, 2.6 km, and 0.9 km correspond to the width or height pixel sizes of 41, 21, 11, 5, 3, and 1 respectively. The prediction accuracy slightly decreases as the size of target area decreases, which implies that the amount of clouds depend on the specific target area and its size.

### C. Prediction for future irradiance

Table V summarizes the prediction results for the irradiance. The proposed SVM-based prediction model yields the best performance among the prediction models in terms of RMSE, MRE, and  $R^2$  values.

The RMSE, MRE, and  $R^2$  values are obtained as 44.1390, 7.7329, and 0.9420, respectively, when  $t = 60$  minutes and  $w_{\text{search}} = h_{\text{search}} = 41$  pixels. Especially, the second best prediction model is the persistence-based model, which was the worst model in the prediction for the amount of clouds.

Figs. 10 and 11 show the measured and predicted irradiance values and the scatter plot of the measured and predicted irradiance, respectively, based on the proposed SVM-based prediction model. Even though some prediction errors occur, the prediction results of the proposed model generally agree with the measured values, and thus, the scatter plot well converges to the red line, which represents the high prediction accuracy of the proposed SVM-based prediction model for future irradiance.

The distribution of prediction errors approximately follows the non-standardized Student's  $t$ -distribution, and the  $\mu$ ,  $\sigma$ , and  $\nu$  values of prediction errors for the irradiance with a prediction horizon of 60 minutes are -1.3945, 34.9578, and 3.9295, respectively.

Table VI and Table VII summarize the prediction results of irradiance based on the proposed SVM-based model for varying the prediction time horizons and radii of the target area, respectively. The irradiance of 15 ~ 150 minutes horizons can be well predicted with  $R^2 = 0.9770 \sim 0.8641$ , while the prediction accuracy degrades from the 180 minutes horizon to  $R^2 = 0.8338$ . In addition, the irradiance prediction of  $t = 60$  minutes shows significant prediction accuracy of  $R^2 = 0.9420$ , 0.9273, and 0.9042 in the range of 35 km ~ 9.4km radii of the target area in Table VII. However, the prediction accuracy for irradiance on smaller target areas (4.3 km ~ 0.9 km radii) slightly decreases to  $R^2 = 0.8596 \sim 0.8277$ .

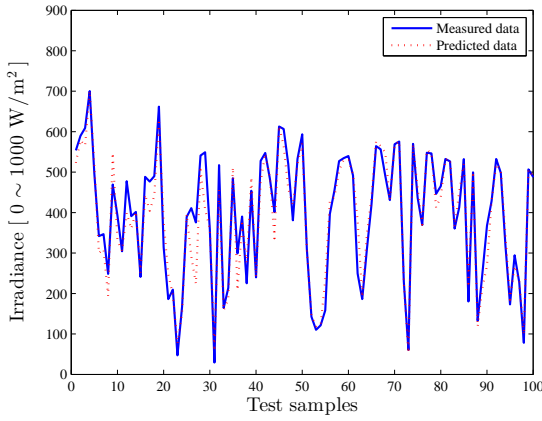


Fig. 10. The measured and predicted irradiance based on the proposed SVM-based model with  $t = 60$  minutes

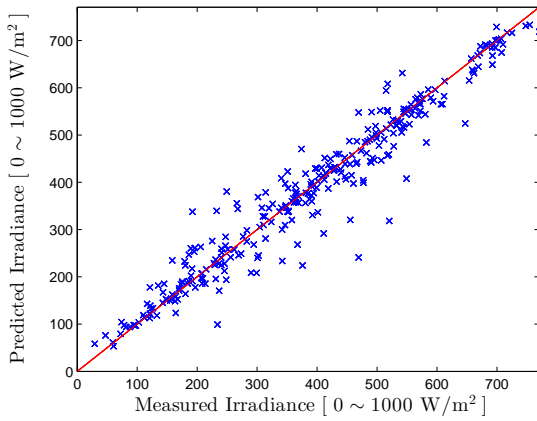


Fig. 11. Scatter plot of the measured and predicted irradiance based on the proposed SVM-based model with  $t = 60$  minutes

TABLE VI

PREDICTION RESULTS FOR IRRADIANCE WITH VARIOUS PREDICTION TIME HORIZONS AND  $w_{\text{target}} = h_{\text{target}} = 41$  PIXELS

Horizons \ Metrics	RMSE	MRE (%)	$R^2$ (0~1)
15 minutes	28.2970	4.6743	0.9792
30 minutes	33.9877	5.6679	0.9670
60 minutes	44.1390	7.7329	0.9420
90 minutes	56.0455	9.3932	0.9155
120 minutes	62.3437	10.9566	0.8830
150 minutes	70.2559	12.0128	0.8641
180 minutes	74.7676	13.9665	0.8338
240 minutes	76.2283	16.2605	0.7906
300 minutes	94.8836	17.9500	0.7309

TABLE VII

PREDICTION RESULTS FOR IRRADIANCE WITH VARIOUS RADII OF THE TARGET AREA AND  $t = 60$  MINUTES

Radius \ Metrics	RMSE	MRE (%)	$R^2$ (0~1)
35 km	44.1390	7.7329	0.9420
18 km	50.2472	8.4920	0.9273
9.4 km	56.0706	9.3361	0.9042
4.3 km	69.9710	12.4461	0.8596
2.6 km	74.7278	13.1583	0.8505
0.9 km	82.3364	14.2098	0.8277

## VII. CONCLUSION

We proposed a solar power prediction model based on a large amount of various historical satellite images and an SVM learning scheme. We thoroughly analyzed 4 years' historical satellite images and utilized them to configure a large number of input and output data sets for the SVM learning. The proposed SVM-based prediction model can simultaneously predict both the future amount of clouds and solar irradiance on multiple sites of South Korea in the range of 15 to 300 minutes (intraday). Throughout the experiments, the proposed SVM-based prediction model shows the highest prediction accuracy, compared to other prediction models such as the conventional time-series and the ANN models. As a result, these prediction data can be effectively utilized for grid operation (load following) and energy management system (EMS) in smart grid.

## ACKNOWLEDGMENT

The authors would like to thank the Korea National Meteorological Satellite Center for providing the high quality satellite images.

## REFERENCES

- [1] "Renewable energy statistics 2013," Korea Ministry of Trade, Industry and Energy, November, 2014, <http://www.motie.go.kr>.
- [2] "List of photovoltaic power stations larger than 50 megawatts (MW)," [http://en.wikipedia.org/wiki/List\\_of\\_photovoltaic\\_power\\_stations](http://en.wikipedia.org/wiki/List_of_photovoltaic_power_stations).
- [3] J. Marcos, L. Marroyo, E. Lorenzo, D. Alvira, and E. Izco, "Power output fluctuations in large scale PV plants: one year observations with one second resolution and a derived analytic model," *Progress in Photovoltaics: Research and Applications*, vol. 19, no. 2, pp. 218–227, 2011.
- [4] J. Kleissl, *Solar energy forecasting and resource assessment*. Academic Press, 2013.
- [5] S. Pelland, J. Remund, J. Kleissl, T. Oozeki, and K. De Brabandere, "Photovoltaic and solar forecasting: state of the art," *IEA PVPS, Task*, vol. 14, 2013.
- [6] A. Hammer, D. Heinemann, E. Lorenz, and B. Lückehe, "Short-term forecasting of solar radiation: a statistical approach using satellite data," *Solar Energy*, vol. 67, no. 1, pp. 139–150, 1999.
- [7] A. Hammer, D. Heinemann, C. Hoyer, R. Kuhlemann, E. Lorenz, R. Müller, and H. G. Beyer, "Solar energy assessment using remote sensing technologies," *Remote Sensing of Environment*, vol. 86, no. 3, pp. 423–432, 2003.
- [8] Z. Peng, S. Yoo, D. Yu, and D. Huang, "Solar irradiance forecast system based on geostationary satellite," in *proc. of IEEE International Conference on Smart Grid Communications (SmartGridComm)*, 2013, pp. 708–713.
- [9] H. Breitkreuz, M. Schroedter-Homscheidt, T. Holzer-Popp, and S. Dech, "Short-range direct and diffuse irradiance forecasts for solar energy applications based on aerosol chemical transport and numerical weather modeling," *Journal of Applied Meteorology and Climatology*, vol. 48, no. 9, pp. 1766–1779, 2009.
- [10] Y. Kaufman, D. Tanré, H. Gordon, T. Nakajima, J. Lenoble, R. Frouin, H. Grassl, B. Herman, M. King, and P. Teillet, "Passive remote sensing of tropospheric aerosol and atmospheric correction for the aerosol effect," *Journal of Geophysical Research: Atmospheres (1984-2012)*, vol. 102, no. D14, pp. 16815–16830, 1997.
- [11] H. G. Beyer, C. Costanzo, and D. Heinemann, "Modifications of the heliosat procedure for irradiance estimates from satellite images," *Solar Energy*, vol. 56, no. 3, pp. 207–212, 1996.
- [12] E. Lorenz, A. Hammer, and D. Heinemann, "Short term forecasting of solar radiation based on satellite data," in *EUROSUN2004 (ISES Europe Solar Congress)*, 2004, pp. 841–848.
- [13] R. Marquez and C. F. Coimbra, "Intra-hour DNI forecasting based on cloud tracking image analysis," *Solar Energy*, vol. 91, pp. 327–336, 2013.



- [14] J. G. Silva Fonseca, T. Oozeki, T. Takashima, G. Koshimizu, Y. Uchida, and K. Ogimoto, "Use of support vector regression and numerically predicted cloudiness to forecast power output of a photovoltaic power plant in kitakyushu, japan," *Progress in Photovoltaics: Research and Applications*, vol. 20, no. 7, pp. 874–882, 2012.
- [15] H. Huang, J. Xu, Z. Peng, S. Yoo, D. Yu, D. Huang, and H. Qin, "Cloud motion estimation for short term solar irradiation prediction," in *proc. of IEEE International Conference on Smart Grid Communications (SmartGridComm)*, 2013, pp. 696–701.
- [16] T. M. Hamill and T. Nehrkorn, "A short-term cloud forecast scheme using cross correlations," *Weather and Forecasting*, vol. 8, no. 4, pp. 401–411, 1993.
- [17] C. W. Chow, B. Urquhart, M. Lave, A. Dominguez, J. Kleissl, J. Shields, and B. Washom, "Intra-hour forecasting with a total sky imager at the UC San Diego solar energy testbed," *Solar Energy*, vol. 85, no. 11, pp. 2881–2893, 2011.
- [18] L. Mora-Lopez and M. Sidrach-de Cardona, "Multiplicative ARMA models to generate hourly series of global irradiation," *Solar Energy*, vol. 63, no. 5, pp. 283–291, 1998.
- [19] W. Ji and K. C. Chee, "Prediction of hourly solar radiation using a novel hybrid model of ARMA and TDNN," *Solar Energy*, vol. 85, no. 5, pp. 808–817, 2011.
- [20] C. Yang, A. Thatte, and L. Xie, "Multitime-scale data-driven spatio-temporal forecast of photovoltaic generation," *IEEE Transactions on Sustainable Energy*, vol. 6, no. 1, pp. 104–112, Jan 2015.
- [21] 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, 2011.
- [22] 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 Transactions on Industry Applications*, vol. 48, no. 3, pp. 1064–1069, 2012.
- [23] H.-T. Yang, C.-M. Huang, Y.-C. Huang, and Y.-S. Huang, "A weather-based hybrid method for 1-day ahead hourly forecasting of PV power output," *IEEE Transactions on Sustainable Energy*, vol. 5, no. 3, pp. 917–926, 2014.
- [24] J.-L. Chen, H.-B. Liu, W. Wu, and D.-T. Xie, "Estimation of monthly solar radiation from measured temperatures using support vector machines—a case study," *Renewable Energy*, vol. 36, no. 1, pp. 413–420, 2011.
- [25] Z. Ramedani, M. Omid, A. Keyhani, S. Shamshirband, and B. Khoshnevisan, "Potential of radial basis function based support vector regression for global solar radiation prediction," *Renewable and Sustainable Energy Reviews*, vol. 39, pp. 1005–1011, 2014.
- [26] B. B. Ekici, "A least squares support vector machine model for prediction of the next day solar insolation for effective use of pv systems," *Measurement*, vol. 50, pp. 255–262, 2014.
- [27] "Korea National Meteorological Satellite Center," <http://nmssc.kma.go.kr/>.
- [28] K. Holmlund, "The utilization of statistical properties of satellite-derived atmospheric motion vectors to derive quality indicators," *Weather and Forecasting*, vol. 13, no. 4, pp. 1093–1104, 1998.
- [29] C. Paoli, C. Voyant, M. Muselli, and M.-L. Nivet, "Forecasting of preprocessed daily solar radiation time series using neural networks," *Solar Energy*, vol. 84, no. 12, pp. 2146–2160, 2010.
- [30] Y. Riffonneau, S. Bacha, F. Barruel, and S. Ploix, "Optimal power flow management for grid connected PV systems with batteries," *IEEE Transactions on Sustainable Energy*, vol. 2, no. 3, pp. 309–320, 2011.
- [31] V. Vapnik, *The Nature of Statistical Learning Theory*. Springer, 2000.
- [32] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and computing*, vol. 14, no. 3, pp. 199–222, 2004.
- [33] D. Dumortier, "Modelling global and diffuse horizontal irradiances under cloudless skies with different turbidities," *Daylight II, jou2-ct92-0144, final report vol.*, vol. 2, 1995.
- [34] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 3, p. 27, 2011.
- [35] H. S. Jang, K. Y. Bae, H.-S. Park, and D. K. Sung, "Effect of aggregation for multi-site photovoltaic (PV) farms," in *proc. of IEEE International Conference on Smart Grid Communications (SmartGridComm)*, 2015, pp. –.
- [36] S. Jackman, *Bayesian analysis for the social sciences*. John Wiley & Sons, 2009, vol. 846.



**Han Seung Jang** received the B.S degree in Electronics and Computer Engineering from Chonnam National University, Gwangju, Korea, in 2012, and the M.S. degree in Electrical Engineering from Korea Advanced Institute for Science and Technology (KAIST), Daejeon, Korea in 2014. Since Mar. 2014, he has been a Ph.D. student in the School of Electrical Engineering, KAIST. His research interests include renewable energy forecasting, electric vehicle (EV) charging management, and machine-to-machine (M2M) communications.



**Kuk Yeol Bae** received the B.S. degree in Electronics Engineering from Ajou University, Suwon, Korea, in 2010, and the M.S. degree in Electrical Engineering from Korea Advanced Institute for Science and Technology (KAIST), Daejeon, Korea in 2012. Since Feb. 2012, he has been a Ph.D. student in the School of Electrical Engineering, KAIST. His research interests include renewable energy forecasting, energy management systems, electricity markets, and machine-to-machine (M2M) communications.



**Hong-Shik Park** received the B.S. degree from Seoul National University, Seoul, Korea in 1977, and the M.S. and Ph.D. degrees from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea all in Electrical Engineering in 1986 and 1995, respectively. In 1977 he joined Electronics and Telecommunications Research Institute (ETRI) and had been engaged in development of the TDX digital switching system family including TDX-1, TDX-1A, TDX-1B, TDX-10, and ATM switching systems. In 1998 he moved to Information and Communications Univ., Daejeon, Korea as a faculty. Currently he is a professor of the School of Electrical and Electronics Engineering, KAIST, Daejeon, Korea. From 2004 to 2012, he was a Director of BcN Engineering Research Center sponsored by KEIT, Korea. His research interests are network architecture and protocols, traffic engineering, and performance analysis of telecommunication systems. He is a member of the IEEE, IEEK and KICS, Korea.



**Dan Keun Sung** received the B.S. degree in Electronics Engineering from Seoul National University in 1975 and the M.S. and Ph.D. degrees in Electrical and Computer Engineering from the University of Texas at Austin, in 1982 and 1986, respectively. Since 1986, he has been with the faculty of the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, where he is currently a Professor with the School of Electrical Engineering. From 1996 to 1999, he was the Director of the Satellite Technology Research Center (SaTReC), KAIST. He had served as Division Editor of the Journal of Communications and Networks from 1998–2007. He also had served as Editor of IEEE Communications Magazine from 2002–2011. His research interests include mobile communication systems and networks, with special interest in resource management, smart grid communication networks, energy networks, machine-to-machine communications, WLANs, WPANs, traffic control in wireless & wired networks, performance and reliability of communication systems, and microsatellites. Dr. Sung is a member of the National Academy of Engineering of Korea. He was the recipient of the 1992 National Order of Merits, the Dongbaek Medal for successfully developing, launching, and operating the first Korean satellite in Korean history, the 1997 Research Achievement Award, the 2000 Academic Excellence Award, the 2004 Scientist of the Month from the Ministry of Science and Technology and the Korea Science and Engineering Foundation, and the 2013 Haedong Academic Grand Award from the Korean Institute of Communications and Information Sciences.