Long-Term Solar Generation Forecasting

Mohana Alanazi, Abdulaziz Alanazi, Amin Khodaei

Department of Electrical and Computer Engineering

University of Denver

Denver, USA

mohana.alanazi@du.edu, alanazi.alanazi@du.edu, amin.khodaei@du.edu

Abstract— **The rapid growth of solar Photovoltaic (PV) technology has been very visible over the past decade. Such increase in the integration of solar generation has brought attention to the forecasting issues. This paper presents a new approach to tackle the long-term forecasting challenge and accordingly reduce the uncertainty of the PV forecast, which would accordingly help facilitate its integration into the electric power grid. The new method includes a set of pre- and post-processes that will be undertaken before the data is fed to the forecasting model and after the forecast is obtained. Using the proposed method, the historical solar PV radiation data, which is non-stationary, is converted to a set of stationary data which will accordingly allow utilization of a larger set of data for forecasting. Numerical simulations exhibit the performance of the proposed method.**

Index Terms—**Photovoltaic, neural network, solar forecasting, global horizontal irradiance (GHI), clear sky irradiance.**

NOMENCLATURE

$GHI_{his.}$	Hourly historical GHI
GHI_{CSK}	Hourly historical clear sky GHI
$GHI_{forecast}$	Forecasted global horizontal irradiance
$GHI_{dev.}$	The difference between $GHI_{his.}$ and GHI_{CSK}
$GHI_{norm.}$	Normalized GHI dataset
$GHI_{denorm.}$	Denormalized GHI dataset
$MAPE$	Mean absolute percentage error
t	Index for year
h	Index for hour
N	Number of samples
w and v	Weights between neurons
k	Index for neurons

I. INTRODUCTION

Renewable energy resources are becoming critical players in the electricity generation sector, primarily due to viability in combating global warming, effectiveness in reducing pollution caused fossil fuel based generation, and diversifying energy mix to ensure energy security and sustainability. Solar energy is one of the most common types of renewable energy that has grown rapidly over the past decade and is anticipated to grow even faster in the future. At the end of 2011, a total capacity of 65 GW solar PV was installed globally, compared to 1.5 GW installed PV capacity in 2000. Over half of that capacity was installed in Germany and Italy, followed by Japan, Spain, the United States and China [1]. In 2013, the cumulative installation capacity of PV systems reached up to 138.9 GW with new installation of 38.4

GW [2], with a large share of market, around 56%, in Asia. The highest PV installation was in China which installed 11.8 GW of PV in 2013, followed by Japan (6.9 GW) and the U.S. (4.8 GW). [2]. In 2013, the U.S. achieved a remarkable trend in solar power. The PV installation increased by 41% in 2013 compared to 2012. This increase resulted in solar power taking the second place in generating electricity, coming only after natural gas. More than half of 2013 installations in the U.S. occurred in California with an installation capacity around 2.8 GW. Arizona was in the second place with a total installation capacity around 0.7 GW. There is now a total of 12.1 GW of PV and 918 MW of concentrated solar plant (CSP) operating in the U.S. [3]. Fig. 1 depicts the percentage share of new generation capacity in the U.S. for the past three years which clearly demonstrates the growing interest in solar generation.
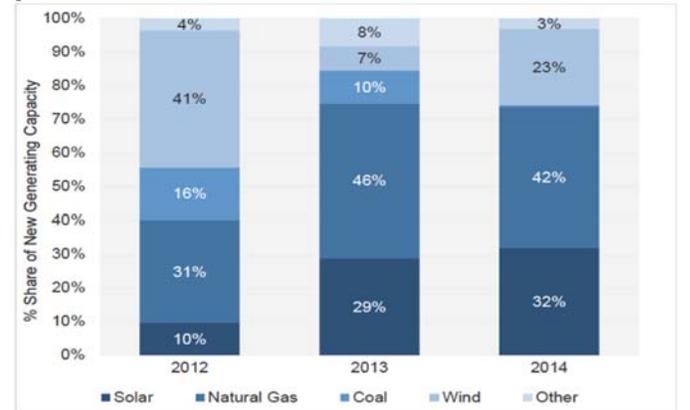
Fig. 1 U.S generation capacity mix for 2012, 2013, and 2014 [4].

Although significantly increasing in the size and installed capacity, the variability and the uncertainty of renewable generation challenges an efficient integration of these resources to the power grid. The variability and the uncertainty are inherent characteristics of renewable resources. Variability refers to the intermittent (i.e., not always available) and fluctuating (i.e., constantly changing from seconds to minutes to hours) nature of renewable generation, while the uncertainty represents the inability to predict in advance the timing and the magnitude of the generation variability. Forecasting aims to accurately determine the future generation values of renewable resource and accordingly reduce the generation uncertainty so that the grid operator will be able to accommodate its variability [5]. A variety of techniques for renewable forecast is proposed in the literature, considering different forecasting horizons and offering different levels of accuracy [6]–[10]. For example in

[11] a MAPE of 16.83% to predict the solar power output for four days is achieved using the recurrent neural network. In [12], the solar power output was forecasted for 24 hours ahead based on weather type classifications (sunny, rainy, or cloudy) using the neural network where a MAPE ranging from 8.29% to 54.44% is achieved.

This paper aims to address a novel forecasting model that helps reducing the uncertainty of the long term PV generation forecasting. The paper presents a comparison between the new proposed method and available methods in long term forecasting and shows that the forecasting error is significantly reduced. The rest of the paper is organized as follows. Section II provides a literature review on existing forecasting methods and challenges. Sections III present the forecasting model outline and formulation of the proposed forecasting method. Section IV provides a comparison of forecasting accuracy between some of the existing methods and the proposed method. The discussions and conclusion are provided in Sections V and VI, respectively.

II.   SOLAR GENERATION FORECASTING CHALLENCHES

Forecasting is not only essential for variable generation, but also useful in load forecasting. In addition, some energy economics quantities such as the electricity price should be forecasted to help with grid's operation, maintenance, and planning [13]. An accurate renewable generation forecast will provide benefits by (i) Minimizing penalties and charges due to imbalance of generated power, (ii) Providing a good knowledge of future energy market trading, and (iii) Helping to carry out reliable operation and maintenance planning [14].

A. Solar Irradiance Forecasting Challenges

The current forecasting methods have confronted variety of challenges that are the source of high forecasting error. When comparing the load forecasting errors to solar irradiance forecasting errors, it is clear that the solar irradiance forecasting is less accurate due to several reasons: First, the time series of solar irradiance is less predictable compared to load forecasting. This is because of the non-stationarity nature of the solar data. The non-continuity of the solar data pattern due to weather changes has imposed significant limitations to forecasting models [15]. So, during the clear sky conditions (sunny) as in Fig. 2a, it is obvious that the patterns are noticeable and the forecasting error is less in these conditions. However, if the weather conditions vary, the pattern of the time series is hardly predictable as in Fig. 2b.

The second challenge that imposes limitations to forecasting model is the change of daytime hours from one day to another during the forecasting horizon (i.e., the sunset and the sunrise). As shown in Table I, the daytime hours change every day. This will impact the pattern of the time series and hence increase the error relatively. The third challenge is the lack of long-term historical solar irradiance data. The long-term forecasting usually requires a long range of dataset to be trained so the model can extract patterns of the time series.

(a)

(b)
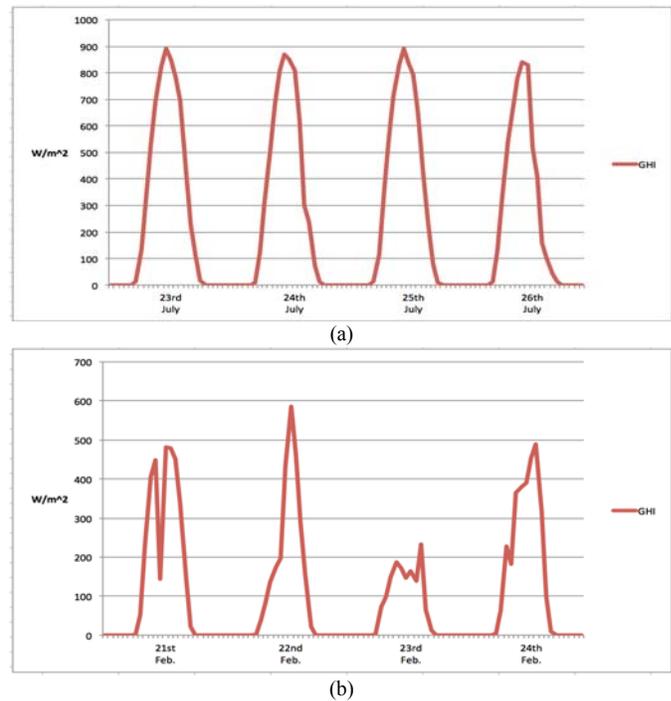
Fig. 2 Solar irradiance for four consecutive days (a) Sunny, (b) Partly cloudy.

Therefore, the solar irradiance is highly dependent on cloud cover and the daytime hours. In addition, the solar irradiance shows a weak stationarity character in terms of a repetitive pattern. Such variations in pattern will cause difficulty to predict any future changes in solar irradiance and limits the application of historical data to highly-correlated ones. Also, the forecasting model requires a large historical dataset to be trained and such amount of data is most likely difficult to be found as solar measurements are limited [15].

TABLE I
SAMPLE OF SELECTED SUNRISE AND SUNSET TIME AND DAYTIME HOURS

Day	Sunrise	Sunset	Daytime hours
22-Jan	7:15	17:06	9 h, 51 min
18-Jul	5:46	20:24	14 h, 38 min
3-Nov	6:30	16:54	10 h, 24 min

B. PV Forecasting Techniques

Forecasting methods can be categorized into three different methods: Physical, Statistical, and Hybrid [16]. Physical models tend to be good for long term forecasting. Two common physical models are the NWP and the satellite sky imagery. The NWP is based on the physics of the atmosphere which uses current observations of the weather and processes this data to predict the future states using super computers. The satellite and cloud imagery based model is a physical forecasting model that analyzes clouds [17], [18]. Under low sun elevations, low irradiance conditions, and high spatial variability, the errors of satellite and cloud images can increase significantly. In [19] a 17% RMSE for half hour cloud index forecast and 30% RMSE for 2 hours forecast is achieved. The statistical method is a mathematical model that uses historical data to predict future values. It is referred to statistical because it utilizes mathematical equations to identify the patterns and trends. Statistical models can be persistence models or time series models that include auto-

regressive (AR), moving average (MA), or both (ARMA). The persistence model is the simplest way for forecasting which basically predicts the future value, assuming it is the same as the previous value. Time series models are based on the historical data and are defined as a sequence of observations measured over time, such as the hourly, daily, or weekly. It is a stochastic process as observations could be random. Hybrid models merge two forecasting techniques to improve the forecast accuracy. They are also known as combined models. The basic idea of hybrid models is to overcome any deficiency of using an individual model, such as regression models, to take the advantage of each individual model and combine them to reduce forecast errors. For instance, the NWP model can be combined with the ANN by feeding the outputs from the NWP as input to the ANN. In [20] a hybrid model was developed by using the satellite imaging as inputs to ANN.

III. THE PROPOSED FORECASTING MODEL

The global horizontal irradiance (GHI) is the main component considered during the PV generation forecasting. The GHI data in specific locations are publically available in many countries, such as the data provided by the National Renewable Energy Laboratory (NREL) in the U.S. [7], [21]. In addition to the GHI historical data, the clear sky GHI is also needed in the proposed method. The clear sky data represents the maximum GHI that could be received during a clear sky day. The solar irradiance is not variable during a clear sky day. Additional data that are collected for the proposed model include cloud cover, temperature, wind speed, and dew temperature. The weather data is available in [22] which is provided by the National Climatic Data Center (NCDC).

The flowchart of the proposed method is presented in Fig. 3 which shows the three stages of the GHI forecasting process, including data pre-process, forecast, and data post-process. Once the final forecast is obtained using the proposed method, the MAPE is calculated to exhibit if it is acceptable. If not, the dataset will be retrained with changing the forecast structure. These stages are described in more detail in the following:

Stage 1: Data Pre-Process

The data preprocessing is a process that occurs before the data fed into the forecasting tool. The preprocessing includes: removing the offset, removing nighttime GHI data, and normalization. To remove the offset, the historical GHI data is subtracted from the clear sky GHI using (1). The resultant data represents the GHI scattered by cloudiness or other factors as shown in Fig. 4.

This data is a function of time and location and reflects all meteorological data that affects solar irradiance. During the forecasting process, other meteorological data is fed to the model to predict the future GHI.

$$GHI_{dev.}(t,h) = GHI_{CSK}(t,h) - GHI_{his.}(t,h)$$
$$h \notin nighttime\ hours \quad (1)$$

$$GHI_{norm.}(t,h) = GHI_{dev.}(t,h)/GHI_{CSK}(t,h)$$
$$h \notin nighttime\ hours \quad (2)$$
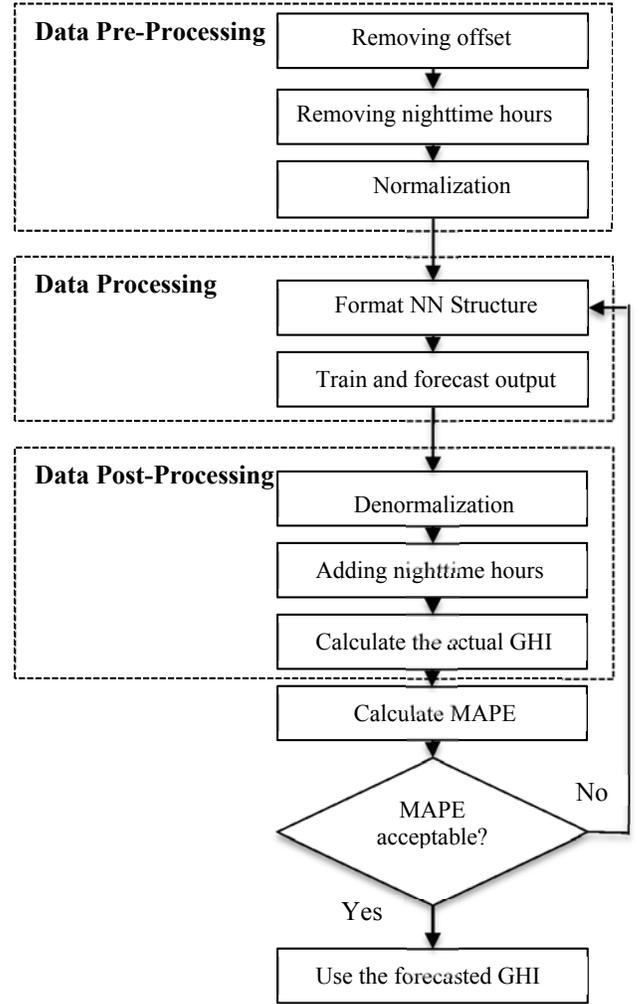
Fig. 3 Proposed flowchart for GHI forecasting process

Once the offsets are removed, the data will be normalized using (2), in which the obtained GHI data is divided by the associated peak clear sky solar irradiance in that day, hence resulting in a value ranging from 0 to 1. Using normalization, all available data will be under the same reference scale, thus the variability as a result of changing solar irradiance peak will be eliminated. The meteorological data is accordingly normalized to [-1,1] using associated daily minimum and maximum values.
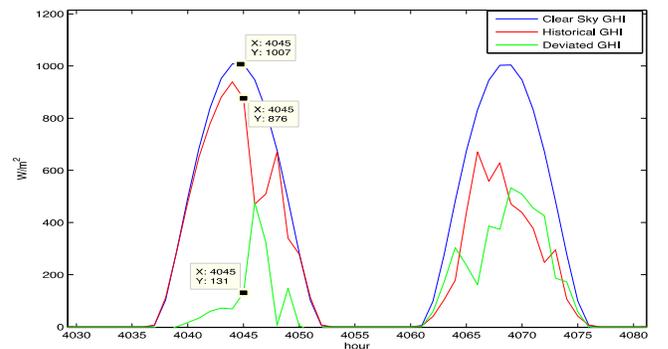
Fig. 4 GHI deviation between clear sky GHI and historical GHI

The last step in data pre-processing is to remove nighttime hours. The solar irradiance varies during the daytime and is zero during the nighttime. However, the daytime hours

change daily as the sunrise and sunset changes. By removing the nighttime hours in this step, only daytime GHI hourly data are remained and will processed. This is accomplished by keeping the daytime hours knowing the exact daily sunrise and sunset as illustrated in Fig. 5. The daytime hours, sunrise, and sunset are fixed for each specific day for the same location over the years. This preprocess is also applied to other input data such as temperature and cloud cover. The complete list of sunrise and sunset times can be found in [23]. An example is provided in Table I.
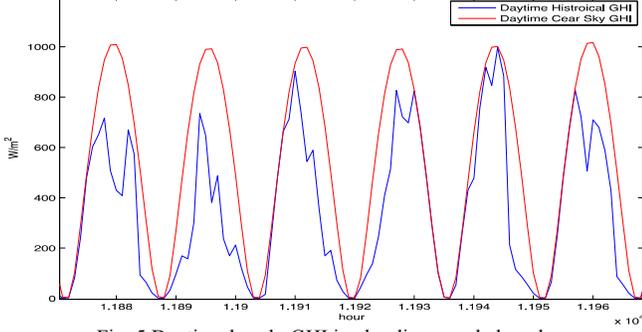
Fig. 5 Daytime hourly GHI in cloudiness and clear day

Stage 2: Forecasting

In the forecasting process, the neural network toolbox is used to predict the future GHI. The GHI data is forecasted in two different ways and the error is compared. The NARX is used to forecast GHI using weather and historical GHI data. The historical and other weather data are fed to the model as input whereas the actual data is fed as a target. The NN structure is established by choosing the number of hidden layers and delays. Once the NN structure is completed, the training process is started. The dataset is trained until the error is minimized between the forecasted and the actual output as illustrated in (3) given the inputs, weights, number of hidden nodes and layers. Here the actual output and the forecasted output are $GHI_{norm.k}$ and $\widehat{GHI_{norm.k}}$ respectively. The weights between neurons are w and v. If the error is not acceptable, the data is retrained after changing the NN structure. The output of the model represents the forecasted GHI data. This paper utilizes NN due to its many benefits in forecasting, however, any other available method can be used without loss of generality of the proposed method.

$$min\, E(w,v,\theta,\gamma) = \frac{1}{N}\sum_{k=1}^{N}\sum_{t=1}^{N1}\sum_{h=1}^{N2}[GHI_{norm.k}(t,h) \\ - \widehat{GHI_{norm.k}}(t,h)]^2 < \varepsilon_1 \tag{3}$$

Stage 3: Data Post-Processing

The forecasted data represents only the daytime hours, and it is normalized. Therefore, three different steps are added after GHI forecasting: denormalization, adding nighttime hours, and calculating the forecasted GHI.

The processed data after forecasting will be multiplied by the daily peak clear sky GHI data to produce the deviated GHI forecasted data using (4). The resulted dataset represents the data that are obtained in the first step in the flowchart. The nighttime hours are further added in a second step to have complete data for each forecasting day. The sunrise and

sunset times in addition to the daytime duration are used to add the removed hours. In the last step, the forecasted data, after denormalization and addition of nighttime hours, is subtracted from the clear sky GHI to obtain the actual forecast values using (5).

IV. NUMERICAL STUDIES

In this section, fifteen years of GHI and clear sky solar data, for the Denver international airport has been used. The dataset represents the interval from 1996 to 2010 which is provided by NREL [7]. Four different weather data are

$$GHI_{denorm.}(t,h) = GHI_{NN}(t,h) * GHI_{CSK}(t,h) \\ \forall h \tag{4}$$

$$GHI_{forcast}(t,h) = GHI_{CSK}(t,h) - GHI_{denorm}(t,h) \\ \forall h \tag{5}$$

$$MAPE = \\ \frac{1}{N} * \sum_{t=1}^{N1}\sum_{h=1}^{N2}\left|\frac{GHI_{his.}(t,h) - GHI_{forecast}(t,h)}{GHI_{his.}(t,h)}\right| * 100 \tag{6} \\ \forall h$$

collected for the same site, including cloud cover, temperature, wind speed, and dew temperature, which is provided by the National Climatic Data Center (NCDC) [22]. The following cases are studies:

Case 1: GHI forecasting using exiting forecasting methods
Case 2: GHI forecasting using the new proposed method

Case 1: The GHI was forecasted using the time series NN provided in MATLAB. The dataset was fed to the model without using any data processing and the error was evaluated using (6). The resulted forecasting GHI shows some distorting values that exceed the actual GHI values. Figure 6a depicts the difference between the actual and the forecasted GHI for the last quarter of the forecasting year. These distortions could be less during the summer times when the weather conditions have less influence in the predicted GHI. One of error sources is the large size of non-stationary dataset that has been used in the forecasting process which results in accumulating errors as the forecasting horizon increases.

Case 2: The proposed method ensures that the dataset is converted from non-stationary to stationary to allow the utilization of a large dataset. First, the dataset is exposed to a pre-process stage where the data are normalized, GHI nighttime values and offset are removed. Then, the dataset is introduced to NN tool where the output GHI is forecasted. Finally, the dataset is exposed to the post-process where resulted GHI data from previous step are denormalized, nighttime values are added, and the GHI data is calculated. Using (6) the MAPE is calculated. In case of a high error, the forecasting tool is adjusted to get a better forecasting performance. Figure 6b shows the actual and forecasted GHI data where both values almost overlap. Forecasting error results in Cases 1 and 2 are listed in Table II.

TABLE II
THE CALCULATED MAPE FOR CASES 1 AND 2

Data Forecasted	MAPE (%)
GHI using available methods	63.78 %
GHI with using the proposed method	9.97%

It is obvious that the proposed method has significantly decreased the MAPE (by almost 84%). Such enhancement in the forecasting performance allows the grid operator to perform a better operation and long term planning. The grid operator will also be able to perform reliable and safe maintenance planning and avoid any risks due to imbalance in power supply.
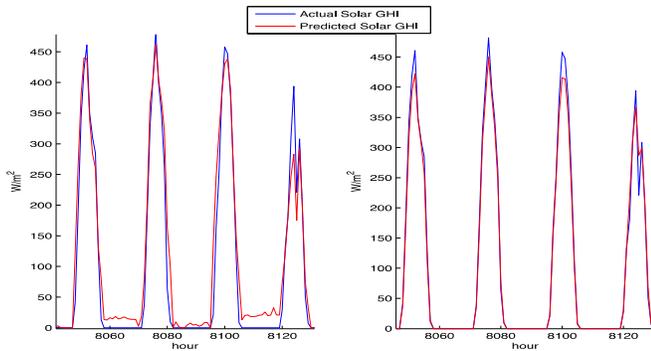
Fig. 6 The hourly GHI for selected days (a) using an available forecasting method, (b) using the proposed forecasting method.

V. DISCUSSIONS

The proposed forecasting model has shown a significant decrease in the forecasting error comparing to the first case where the dataset is fed to the model without any processing. The new model has many features compared to the existing methods:

- The model converts the non-stationary data to stationary, where the statistical properties of the given time series, such as mean and variance, are constant functions of time.
- The proposed model reduces the dataset size almost to half because it's not including the nighttime values when the GHI is zero. This will accelerate the time needed to perform the forecasting.
- The accuracy of the forecasting is significantly improved, hence the grid operator will be able to carry out reliable operation, maintenance, and planning.

VI. CONCLUSIONS

The accuracy of the global horizontal irradiance forecast is an important factor due to many operational, economical, and engineering reasons. The objective of this work was to propose a novel long-term forecasting method which was based on data pre- and post-process. The novelty of this study was to analyze the dataset before it was fed to the forecasting model in order to make sure that the dataset is stationary. The processed time series data was fed into a neural network allied with the nonlinear autoregression with external input (NARX). The proposed method showed an improvement in the accuracy of the forecasting, where the long-term MAPE was substantially reduced from 63.78% to 9.97%. This promising result can be further applied to shorter time horizons, such as day-ahead forecasting, which will be considered as future work of this research.

REFERENCES

[1] "IEA - Solar (PV and CSP)." [Online]. Available: http://www.iea.org/topics/solarpvandcsp/. [Accessed: 20-Aug-2014].

[2] "GLOBAL MARKET OUTLOOK For Photovoltaics," European Photovoltaic Industry Association (EPIA), 2013. [Online]. Available: http://www.epia.org/news/publications/global-market-outlook-for-photovoltaics-2014-2018/.

[3] "Solar Market Insight Report 2013 Year in Review | SEIA." [Online]. Available: http://www.seia.org/research-resources/solar-market-insight-report-2013-year-review. [Accessed: 02-Sep-2014].

[4] "Solar Market Insight Report 2014 Year in Review | SEIA." [Online]. Available: http://www.seia.org/research-resources/solar-market-insight-report-2014-q4. [Accessed: 14-May-2015].

[5] G. M. Masters, Renewable and efficient electric power systems. Hoboken, NJ: John Wiley & Sons, 2004.

[6] A. Mellit and A. M. Pavan, "A 24-h forecast of solar irradiance using artificial neural network: Application for performance prediction of a grid-connected PV plant at Trieste, Italy," Sol. Energy, vol. 84, no. 5, pp. 807–821, May 2010.

[7] Y. Huang, J. Lu, C. Liu, X. Xu, W. Wang, and X. Zhou, "Comparative study of power forecasting methods for PV stations," in 2010 International Conference on Power System Technology, 2010, pp. 24–28.

[8] Y. Zhang, M. Beaudin, R. Taheri, H. Zareipour, and D. Wood, "Day-Ahead Power Output Forecasting for Small-Scale Solar Photovoltaic Electricity Generators," IEEE Trans. Smart Grid, pp. 1–1, 2015.

[9] R. Marquez and C. F. M. Coimbra, "Forecasting of global and direct solar irradiance using stochastic learning methods, ground experiments and the NWS database," Sol. Energy, vol. 85, no. 5, pp. 746–756, May 2011.

[10] K. Stefferud, J. Kleissl, and J. Schoene, "Solar forecasting and variability analyses using sky camera cloud detection & motion vectors," in Power and Energy Society General Meeting, 2012 IEEE, 2012, pp. 1–6.

[11] C. Chupong and B. Plangklang, "Forecasting power output of PV grid connected system in Thailand without using solar radiation measurement," Energy Procedia, vol. 9, pp. 230–237, 2011.

[12] C. Chen, S. Duan, T. Cai, and B. Liu, "Online 24-h solar power forecasting based on weather type classification using artificial neural network," Sol. Energy, vol. 85, no. 11, pp. 2856–2870, Nov. 2011.

[13] V. P. Singh, K. Vaibhav, and D. K. Chaturvedi, "Solar power forecasting modeling using soft computing approach," in Engineering (NUiCONE), 2012 Nirma University International Conference on, 2012, pp. 1–5.

[14] "The importance of wind forecasting - Renewable Energy Focus." [Online]. Available: http://www.renewableenergyfocus.com/view/1379/the-importance-of-wind-forecasting/. [Accessed: 05-Sep-2014].

[15] Y. Zhang, M. Beaudin, R. Taheri, H. Zareipour, and D. Wood, "Day-Ahead Power Output Forecasting for Small-Scale Solar Photovoltaic Electricity Generators," IEEE Trans. Smart Grid, pp. 1–1, 2015.

[16] H. M. Diagne, M. David, P. Lauret, and J. Boland, "Solar irradiation forecasting: state-of-the-art and proposition for future developments for small-scale insular grids," American Solar Energy Society, 2012.

[17] S. Cros, O. Liandrat, N. Sébastien, and N. Schmutz, "Extracting cloud motion vectors from satellite images for solar power forecasting," in Geoscience and Remote Sensing Symposium (IGARSS), 2014 IEEE International, 2014, pp. 4123–4126.

[18] 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.

[19] A. Hammer, D. Heinemann, E. Lorenz, and B. Lückehe, "Short-term forecasting of solar radiation: a statistical approach using satellite data," Sol. Energy, vol. 67, no. 1, pp. 139–150, 1999.

[20] R. Marquez, H. T. C. Pedro, and C. F. M. Coimbra, "Hybrid solar forecasting method uses satellite imaging and ground telemetry as inputs to ANNs," Sol. Energy, vol. 92, pp. 176–188, Jun. 2013.

[21] "Solar Data 1991-2010 Site #725650." [Online]. Available: http://rredc.nrel.gov/solar/old_data/nsrdb/1991-2010/hourly/siteonthefly.cgi?id=725650. [Accessed: 16-Sep-2014].

[22] "Beautiful Weather Graphs and Maps - WeatherSpark." [Online]. Available: http://weatherspark.com/. [Accessed: 16-Sep-2014].

[23] "Denver International Airport, CO, United States Of America Sunrise and Sunset times - weather.co.uk." [Online]. Available: http://uk.weather.com/climate/sunRiseSunSet-Denver-International-Airport-CO-USCO0425?month=1. [Accessed: 17-Sep-2014].