21, rue d'Artois, F-75008 PARIS
http : //www.cigre.org

CIGRE US National Committee
2016 Grid of the Future Symposium

Investigating the Impact of Time Series Stationarization on Day-Ahead Solar Forecasting

M. ALANAZI, A. KHODAEI
University of Denver
USA

S. BAHRAMIRAD, E.A. PAASO
ComEd
USA

SUMMARY

In some geographic regions solar energy sources have become attractive generation alternatives for grid planners and customers in recent years. An efficient solar forecasting method, which takes into account generation variability and is able to identify associated uncertainty, helps alleviate many of the integration challenges and ensures a reliable, safe, and cost-effective deployment. Solar forecasting is a decisive factor that supports grid management, contributes to stability in electricity market, and helps perform reliable power operations. This paper focuses on the suitability of the input data sets for solar forecasting and investigates how changing the state of the solar data set from non-stationary to stationary would affect forecast results through numerical simulations. The augmented Dickey–Fuller (ADF) test is applied to validate the stationary state of the data set before performing the forecasts and making comparisons. The results will show that a mean absolute percentage error (MAPE) of less than 1% can be achieved using the stationary data set under different weather conditions with reductions of up to 70% compared to the case without using the proposed model.

KEYWORDS

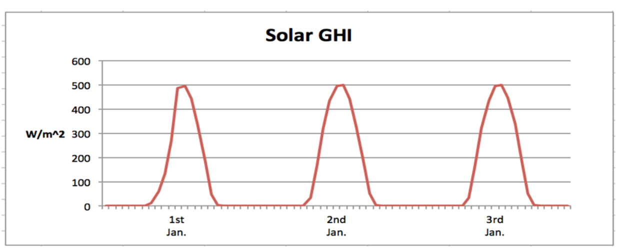
Solar photovoltaic, neural network, global horizontal irradiance (GHI).

Mohana.Alanazi@du.edu

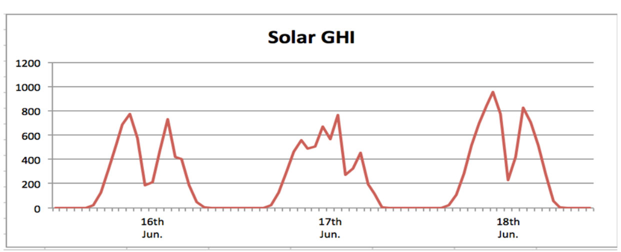
1. INTRODUCTION

The boost of energy supply from variable generation resources, particularly from wind and solar, has increased in recent years. Renewable energy sources can provide sustainable generation alternatives for fossil fuel based power supply. This major shift from fossil fuel based generation is due to the environmental concerns related with CO2 emissions and global warming. However, renewable generation sources have created operational challenges to the electric grid due to their output variability uncertainty. At high levels of penetration, these resources may adversely affect the reliability and power quality of the power grid. Sudden variations in the power output of solar and wind may cause an inimical effect on power system operation. As a result, the deployment of renewable resources has encountered different challenges. It is generally known, however, that by utilizing a highly accurate forecast many of these challenges can be efficiently addressed, enabling grid operators to effectively plan ahead for managing generation variabilities. Thus, solar power forecast is of utmost importance to address variability and uncertainty of solar power output [1].

There are a few major sources of error in solar forecasting: (i) the time series of solar irradiance is unpredictable, caused mainly by weather changes and partial/full cloud cover. As a result, the solar time series is considered non-stationary. Figure (1a) depicts a clear pattern of hourly solar irradiance in clear sky days while figure (1b) depicts the fluctuations and changes in the patterns due to climate changes. Statistical methods, such as learning-based models used in forecasting, require the time series to be stationary; and (ii) the solar irradiance changes every day based on the duration of the day and sunrise/sunset times. When the duration is different, the historical data cannot be easily used to forecast solar irradiance. For example, the solar data from one day before, one week before, or one month before cannot be suitable to make a viable forecast, while the data from one year before (on the same exact date) is useful as it has similar sunrise/sunset times. This drawback limits the number of available data points to a set of selected points, which may not be adequate to perform an accurate forecast [2, 3].



(a)



(b)

Fig. 1 Solar GHI in Denver for 2013 in (a) Sunny, (b) cloudy days

This paper builds on the previous work of authors [4] in developing novel solar forecasting methods and further investigates how a set of stationary data can improve the forecasting results. The stationary state can be achieved by using different techniques, such as differencing and detrending. The rest of the paper is organized as follows: Section 2 presents the proposed solar forecasting method and shows how non-stationary data can be converted to stationary data; Section 3 presents simulation results on test data to show the performance of the converted data; and Section 4 provides conclusions for the paper.

2. PROPOSED FORECASTING MODEL

The proposed model by authors in [4] considers the global horizontal irradiance (GHI) for forecasting purposes, which is the total irradiance received at the surface and consists of both Direct Normal Irradiance (DNI) and Diffuse Horizontal Irradiance (DHI). The global horizontal irradiance is provided by National Renewable Energy Laboratory (NREL) and is available to the public in [5]. The historical GHI and clear sky GHI, which is the maximum GHI received at the surface during clear sky conditions, are collected and used. The model includes three stages of data pre-processing, forecasting, and data post-processing as shown in Figure 2.

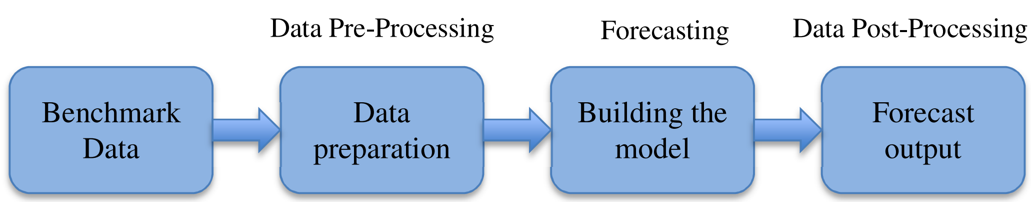


Fig. 2 Flow diagram for solar irradiance forecasting

Stage 1 Data Pre-Process: After the historical data set is gathered, it is sent to the data pre-processing stage. This stage will first remove the offset, from historical GHI, by subtracting the historical GHI from the clear sky GHI as in (1). The resultant GHI from (1) represents the scattered GHI by cloudiness and other factor as shown in Figure 1.

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

Next, the nighttime hours are removed as the solar irradiance values at nighttime are zero. The nighttime hours are obtained based on the sunrise and sunset times in each day. The resultant time series of solar irradiance from this step is still non-stationary, so the data are introduced to the fitting model in order to detrend the data. There are various available detrending models [6, 7] that discuss how the Al-Sadah model outperforms many others, such as Jain, Baig, and Kaplains. The Al-Sadah's output is presented as follows:

$$I_t = a_0 + a_1 h + a_2 h^2 + \cdots + a_n h^n \qquad h \notin nighttime\ hours \tag{2}$$

where $a_0, a_1 \ldots a_n$ are constants and can be found by the fitting and actual data. The Augmented Dickey–Fuller (ADF) test is used to examine whether the new data set is stationary or not by checking the existence of a unit root. If the unit root exists, the time series is non-stationary and the null hypothesis should be accepted, otherwise, the null should be rejected and the time series is stationary. The ADF test is presented as follows:

$$\Delta x_t = \mu + \beta t + \rho x_{t-1} + \delta_1 \Delta x_{t-1} + \delta_2 \Delta x_{t-2} \ldots + \delta_p \Delta x_{t-p} + e_t \qquad (3)$$

where μ is a constant called drift, β is a coefficient that represents the trend, p is the number of lags or the order of the autoregression process, and e_t represents random variables with zero mean. The last step in data pre-processing is normalization, which ensures that the data sets are under the same reference scale and prevents variability due to solar irradiance peaks. The normalization is performed by dividing the obtained data points by the associated clear sky GHI as in (4).

$$I_{tNorm.}(t,h) = I_{t.}(t,h)/GHI_{CSK}(t,h) \qquad h \notin nighttime\ hours \qquad (4)$$

Stage 2 Forecasting: The resultant data set from the previous stage is stationary and ready to be introduced to the forecasting tool. Neural network is used as the desired forecasting tool, [8], however, it can be replaced with any other forecasting tool based on the operator's discretion. The stationary solar irradiance time series is fed to the tool and the training process is started. The training process is repeated while changing the characteristics of the neural network model, i.e., number of input nodes, number of hidden layers, etc. to find the most suitable configuration that minimizes the forecast error. It is worth mentioning that from the available data set a large percentage (80%) can be used for model training purposes while a small percentage (the rest) can be used for testing and validation.

Stage 3 Data Post-Process: This stage reverses the steps performed in the first stage. The forecasted data from stage 2 represents only daytime values in a normalized form. The three processes conducted in this stage are: de-normalizing the forecasted data (5), adding the night time GHI values, and calculating the actual forecasted GHI by adding the offset (i.e., clear sky GHI) to the forecasted data (6). The obtained GHI from this stage is the actual forecasted GHI that can be used to calculate solar generation.

$$GHI_{denorm.}(t,h) = GHI_{NN}(t,h) * GHI_{CSK}(t,h)\ ,\ \ \forall h \qquad (5)$$

$$GHI_{forcast}(t,h) = GHI_{CSK}(t,h) - GHI_{denorm}(t,h), \forall h \qquad (6)$$

There are various metrics to measure the accuracy of the obtained forecast. The mean absolute percent error (MAPE) as in (7) is proposed here, however, it can be replaced with any other metric to analyse the forecast accuracy.

$$MAPE = \frac{1}{N} * \sum_{t=1}^{N} \left| \frac{GHI(t)_{actual} - GHI(t)_{forcast}}{GHI(t)_{actual}} \right| * 100 \qquad (7)$$

3. CASE STUDY

A day-ahead forecast under various weather conditions is performed to show the significance of the stationary data sets in improving the forecast accuracy. MAPE is calculated to evaluate the performance under each case.

Case 1 Forecast using non-stationary data:

The hourly GHI data for March 2010 is tested, using ADF, for checking stationarity and the output results are summarized in Table I. As presented, the test result is above the critical value, which indicates that there is a unit root and null hypothesis should be accepted. As a result the available solar irradiance is a non-stationary time series. This data is directly fed to the neural network forecasting tool to forecast GHI values for different test dates under various weather conditions. The resultant MAPE is summarized in Table 2. As the obtained results indicate, the forecast is more accurate under clear sky conditions, i.e., a sunny day, compared to other days with cloud cover, conceivably due to the GHI variability during cloudy days. Moreover, the error in cloudy and partly cloudy days is almost the same, but it cannot be generalized for other days with similar weather conditions.

Table 1: The ADF Test For Hourly Average GHI For March 2010 – Case 1

Statistical Power	**Significance level**	**Test result**	**Critical value**
0.47	0.05	-0.465	-1.9567

Table 2: Forecast Performance – Case 1

Weather Condition	**Day**	**MAPE (%)**
Partly Cloudy	April 8	2.624
Cloudy	May 5	2.670
Sunny	August 13	1.117

Case 2: Forecast using stationary data:
The same data set, for March 2010, is used in the proposed forecasting model. The data set is detrended using Al-Sadah's model. Figure 5 depicts the actual hourly average GHI and the calculated fitting model. Then the residual, which represents the difference between the actual and the fitted model, is calculated and introduced to the ADF test. The test result is summarized in Table 3. The test shows that the result is below the critical value and that means there is no unit root and the null hypothesis should be rejected. As a result, the resultant time series is a stationary time series.

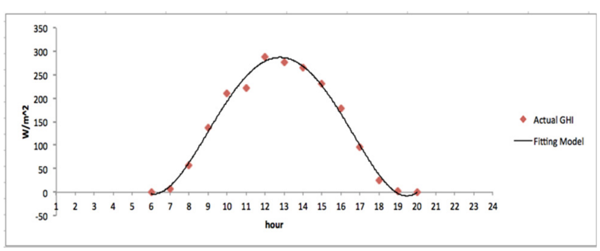


Fig. 3 Actual hourly average and the fitting model series for month of March 2010

The stationary data set is fed to the same forecasting tool and the GHI values for the same days as in Case 1 are calculated. The summary of the evaluation, based on MAPE calculations, is presented in Table 4. As the results show, MAPE values are considerably reduced compared to Case 1, demonstrating a more accurate forecast when using the stationary data set. Moreover, the MAPE values for different days are much closer to each other than in Case 1, which shows that using stationary data will improve forecasts not only under clear sky conditions but also for cloudy days.

Table 3: The ADF Test For Hourly Average GHI For March 2010 – Case 2

Statistical Power	Significance level	Test result	Critical value	RMSE	NRMSE
0.001	0.05	-5.12	-1.957	4.30	0.032

Table 4: Forecast Performance – Case 2

Weather Condition	Day	MAPE (%)
Partly Cloudy	April 8	0.8107
Cloudy	May 5	0.7998
Sunny	August 13	0.6385

4. CONCLUSION

In this paper, a comparison was made between the application of stationary and non-stationary data sets in forecasting day-ahead solar GHI values. The three-stage model, introduced by authors, was utilized to convert the non-stationary data set to a stationary data set and perform the forecasting. Both stationary and non-stationary data sets were fed into a similar forecasting tool, based on neural network, and the resulting forecasts were compared in terms of errors. The results show that the application of stationary data set could reduce the MAPE ranges by as much as 42% in sunny days and 70% in cloudy days. The high accuracy of the forecast would allow power system operators to predict any sudden fluctuations in the solar output and perform the proper control actions.

BIBLIOGRAPHY

[1] A. Tuohy, J. Zack, S. E. Haupt, J. Sharp, M. Ahlstrom, S. Dise, E. Grimit, C. Mohrlen, M. Lange, M. G. Casado, J. Black, M. Marquis, and C. Collier, "Solar Forecasting: Methods, Challenges, and Performance," (IEEE Power Energy Magazine, vol. 13, no. 6, pp. 50–59, Nov. 2015).

[2] D. Dickey and W. Fuller, "Likelihood ratio statistics for autore- gressive time series with a unit root," (Econometrica Vol. 49 No. 4, pp. 1057–1072, 1981).

[3] M. Alanazi, A. Alanazi, and A. Khodaei, "Long-Term Solar Generation Forecasting" (IEEE PES Transmission and Distribution Conference, Dallas, TX, 2016)

[4] M. Alanazi, and A. Khodaei, " Day-Ahead Solar Forecasting Using Time Series Stationarization and Feed-Forward Neural Network" (North American Power Symposium, Denver, CO, 2016)

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

[6] 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, May 2011).

[7] K. Benmouiza and A. Cheknane, "Small-scale solar radiation forecasting using ARMA and nonlinear autoregressive neural network models" (*Theor. Appl. Climatol.*, vol. 124, no. 3–4, pp. 945–958, May 2016).

[8] M. Abuella and B. Chowdhury, "Solar power forecasting using artificial neural networks," (North American Power Symposium, 2015).