A Review Paper on Correction Methods for Solar Irradiances Data

Suzan Katamoura [1], Layla Hajar[2] and Mehmet SabihAksoy[3]

[1]Statistics and Analysis Officer, Studies and Research Sector, King Abdullah City for Atomic and Renewable Energy, Riyadh 11451, Saudi Arabia; s.katamoura@energy.gov.sa

[2]PHD Student, Department of Information Systems, College of Computer and Information Sciences, King Saudi University, Riyadh 11451, Saudi Arabia; lhajr@ksu.edu.sa

[3]Professor, Department of Information Systems, College of Computer and Information Sciences King Saudi University, Riyadh 11451, Saudi Arabia; msaksoy@ksu.edu.sa

ABSTRACT:There is big data related to solar resources and associated weather parameters that are actively collected from more than 45 solar radiation measurement stations operating around Saudi Arabia. King Abdullah City for Atomic and Renewable Energy (K.A.CARE) has been managing this data since 2013 and making it available through Renewable Resources Atlas Portal. This data uses to support stakeholders in decision-making related to renewable energy projects in Saudi Arabia. However, maintaining such a large-scale network of measurement devices and receiving streaming data every minute over the years might result in some data issues in certain periods. These issues can result from different reasons. In this case, the data was slightly affected while maintaining the sensors of measurement devices, particularly the physical cleaning schedule. This paper aims to review available solar data correction methods in the literature to solve these instances and correct data values. The goal is to investigate the different techniques and find the most applicable to implement on K.A.CARE data.

Keywords: Solar Data, BigData,Data Correction

1. Introduction

King Abdullah City for Atomic and Renewable Energy (K.A.CARE) has a solar measurement network of more than 45 stations operating around Saudi Arabia. These ground stations collect data every minute and store them in the K.A.CARE datacenter. Today, there are 335 Gigabytes of stored data, and the database continuously receives more incoming data. This big structured data comprises Solar irradiances and associated weather parameters for eight consecutive years. K.A.CARE is projecting this data through a web portal, Renewable Resources Atlas.

Importance of Data

Solar data is vital for renewable energy projects such as building power plants and solar farms. Also, data is used for solar resources assessment to find the best locations for projects. In addition, data is a key to improve the accuracy of the satellite models that offer solar resource analysis for projects. All that is because the bankability of the projects depends on the accuracy and quality of the solar radiation data collected by the ground stations' measurements.

Furthermore, the data analysis is conducted for insights to assist strategic business decisions making and improve their quality in Saudi Arabia. Therefore, decisions on funding the project at an acceptable and quantifiable risk level are depending on the data accuracy used for analysis [1]. Accordingly, data accuracy is the prime factor representing a degree of data quality considering precision, completeness, and consistency.

However, maintaining this massive volume of data and many different types of measurement devices is not an easy task. Consequently, some data issues occur in certain periods and require corrective action. These issues can result from various reasons. In this case, the data was affected while maintaining the sensors of measurement devices, particularly the physical cleaning schedule. As a result, the dust was accumulated on sensors causing a slight variation to the values obtained. This inaccuracy would affect the performance of concentrated solar power (CSP) and thus the production cost for electricity [2]. Accordingly, it is essential to fix variation in data value to reach the higher possible accuracy level. In this context, the soiled data were identified through the data quality and validation process using different analysis techniques and tools by K.A.CARE team.

This work aims to review available solar data correction methods in the literature for soiled solar data, investigate the different techniques and find the most appropriate approach to correct data values to implement on K.A.CARE data.

Data Correction Concept

Data correction is close in concept to data cleansing/data cleaning. It is the process to identify and address issues in data to improve data quality to reach the highest precision and accuracy of data and to be able to extract useful information to meet the business needs [3]. Naturally, that will influence the information deduced from the data considerably. Generally, data issues include inaccurate, imprecise and incomplete values. These quality issues are associated with real-time data acquisition and complexity and diversity of data [4]. The correction expected for solar data is to restore the actual readings of the measurement sensors that was reduced due to the soil accumulation on the sensor's surface.

Since soiling on the sensors is the primary source for underestimating the measured irradiation, especially at stations where the cleaning schedule was not maintained, this paper will discuss the available methods for correcting solar data in the literature.

Methods for correcting Solar Irradiances Data

Geuder and Quaschning (2006) paper "Soiling of irradiation sensors and methods for soiling correction" discussed soling characteristics and their impact on solar irradiances measurement. The study found that soiling varies between different instruments and depends on location. Also, soiling might increase or decline over time depending on weather conditions like rains or sand storms. Then it recommended methods for soiling correction. The recommendations included daily sensor cleaning, redundant instruments at the same site to compare measured and calculated values and proposed an updated formula with a general soiling parameter to calculate the real value for solar irradiances [5].

Another study by Schwandt et al. (2019), "Soiling Impact on Direct Normal Irradiation Measurements", conducted further examinations at two sites in Spain and Germany over one year. Similar to previous research and concurring with many studies, it was found that soiling rate depends on station location characteristics. Furthermore, it was concluded that daily soiling rates follow an exponential dependency until a certain percentage, then follow a linear behavior over time [6].

Findings from the preceding works agree with K.A.CARE quality system and considerations concerning soiling characteristics and rates behavior. Nevertheless, The cleaning recommendation is a preventive method, and this was the source of the variations that occurred in K.A.CARE data. Similarly, the redundant instruments existed in some stations, yet they are suffering from soling. Also, the formula developed depended on the constant soiling factor, which will not work well in this case because of the large land and very diverse and violent climate conditions.

The work of DeFreitas et al. (2019), “Evaluating the Accuracy of Various Irradiance Models in Detecting Soiling of Irradiance Sensors” evaluated the viability of several models that depends on clear-sky status as well as purchased satellite data to estimate the soiling of a reference irradiance sensor. Approximation using purchased satellite data as a reference yield an accurate result. However, the results were more accurate when local meteorological conditions were considered in the model. Also, the clear-sky model performs better in some cases offering a low-cost tool that detects soiling of irradiance sensors. Additionally, the researchers used a simple formula to compute the soiling ratio [7].

Findings from this work are interesting, and the formula for soiling ratio seems reasonable and easy to implement and will yield accurate corrected values with few considerations. However, five primary concerns with this approach:

(1) worked on one-hour data resolution, while K.A.CARE data is at one-minute resolution.

(2) it considered only clear-sky status while the concerned data is not limited to that condition.

(3) the study used a reference sensor next to the affected sensor as planned, yet the incident in K.A.CARE is actual (unplanned).

(4) the results were imprecise under certain conditions and led to an approximation of soiling rate with higher uncertainty.

(5) Comparing ground data to satellite data as a reference will not be precise as the satellite data overestimates the value of the solar irradiance by 9%.

In the study of “A non-parametric method for correction of global radiation observations”, by Bacher et al. (2013), a technique for alignment of solar irradiances (global radiation) soiled observations, from a weather station, with calculated global radiation values from a forecast model based on physical principles. Similar to work [5], the correction method adopted in this study used a statistical clear-sky model applied on hourly data to detect systematic deviations between the two observations. Then, it corrected solar irradiance values based on corresponding values from the forecasting model. Next, the results were validated against observations of another station. The research found that the correction method can optimize the usage of solar radiation observations for a set of applications depending on this information [8].

The study presented some advantages: the technique works semi-automatically and requires only the observations and location with minor tunes to a few parameters. Furthermore, it is possible to embed solar irradiance data used for correction from any calculated clear-sky model. In contrast, several remarks on this study need to be considered for K.A.CARE required solution. First, it uses the hourly data while one-minute data is targeted by K.A.CARE data correction. Second, it uses a forecasting model for the comparison process, not actual data. Third, it works on a particular solar parameter, global radiation, and must be modified to work separately for other solar irradiance components. Finally, the validation is made against different a station where it could make against data from previous years of the same station or satellite data for the exact location.

Another general correction method that tackled different errors in data, called bias, is in the literature, "Bias Correction Method for Solar Radiation Based On Quantile Mapping to Provide Weather Data for Building Energy Simulations" by Arima (2016). The proposed approach, quantile mapping, was developed to correct the bias of precipitation or temperature, yet some studies applied it for solar radiations correction. For the solar radiation bias correction, only the data average is used for this purpose. The model uses the daily data average under clear-sky conditions. The result of the study showed underestimation for solar radiation by 6% occurred when using corrected data by this method [9].

It is essential to note that the study concept might still apply to the data issue under investigation in K.A.CARE and highlight some considerable points, but using daily average resolution and clear-sky model raises concerns.

One more study by Soulayman et al. (1995), "A Simple Method for Correcting Solar Radiation Measurements Made Using Non-Calibrated Eppley-Type and Robitzsch-Type Pyranometers", proposed a correction method for solar data variations resulting from non-calibrated instruments. The approach is based on a statistical comparison between the distribution of the observations from two sensors; one is non-calibrated, and the second is a calibrated sensor. Then a correction factor is obtained. The study was also built on clear-sky conditions and considered the sensor lifetime behavior and climatic conditions. The work revealed that the proposed method significantly reduced solar data variations, increasing the accuracy of corrected values [10].

Although the work is designed for non-calibration variations, it could be considered for the soiling variation in K.A.CARE data. Yet, it needs to be examined on the soiled data. Also, since there is no unsoiled reference sensor at the exact locations where soiled data was collected, it might be considered to compare data distribution for the same sensor over the same period in historical data where the sensors adhered maintenance schedule.

Muller et al. (2017), in their paper on "A Method to Extract Soiling Loss Data from Soiling Stations with Imperfect Cleaning Schedules", studied a method to determine the soiling effect by comparing the output energy from soiled PV cell of nine stations to that of a maintained PV cell. The algorithm computed the daily soiling ratio between 11 am and 1 pm for the affected periods. In addition, the research examined an automated approach to detect and correct soiling errors. It was proved from initial results that this method reduced soiling variations. However, the corrective algorithm could not handle partially cleanings resulting from melting snow on the surface of PV reference cells because the algorithm considered complete cleaning in its design [11].

The work carried out different data types than solar irradiances, yet there were some suitable lessons. First, the daily soiling ratio of PV cells was calculated, which can be applied to the solar data in the same manner. Also, it focused on daytime readings which make sense for solar irradiances. However, there will be some challenges with K.A.CARE data solution since neither information on the soiling level existed nor a reference for a clean sensor. Also, the daytime in Saudi Arabia starts earlier than 11 am; thus, it needs to be considered when data is used.

Conclusion and Future Work:

This work aims to review the literature and examine the available approaches to correct variations on solar irradiances data. This is vital to increase the benefit of the data for the renewable energy project, solar resources assessment and forecasting, and support Saudi Arabia's stakeholders in the decision-making process.

There were different methods introduced and analyzed. Each approach has advantages and related comments when it comes to implementation on K.A.CARE data. In conclusion, the method to be adopted to solve the variation of K.A.CARE data due to soiling must provide accurate values, and has to consider the following recommendations:

(1) The model should calculate the daily soiling factor based on historical data of the same sensor.

(2) The study should consider the data during daytime; for example, from 7 am to 7 pm.

(3) The data resolution should be at one-minute level

(4) The different weather conditions have to be considered and not only clear-sky.

(5) Select the suitable formula to correct the values using the daily soiling ratio

(6) Use satellite data and historical observations to validate the corrected data.

The next step is to work on building the model to correct data and do the necessary testing and validation to assure the accuracy and quality of the resulting values.

Acknowledgment:

The authors would like to thank Dr. Yousef al Ghamdi for his technical advice in the engineering field, and Eng. Reham Hafiz for her assistance in the data information.

References:

[1] E. Zell, S. Gasim, S. Wilcox, S. Katamoura, T. Stoffel, H. Shibli, J. Engel-Cox, M. Al Subie, “Assessment of solar radiation resources in Saudi Arabia”, Solar Energy, Volume 119, 2015, Pages 422-438.

[2] Cebecauer, Tomas & Suri, Marcel. (2016). Site-adaptation of satellite-based DNI and GHI time series: Overview and SolarGIS approach. 1734. 150002. 10.1063/1.4949234.

[3] Sardjono, Sardjono&Alamsyah, R. &Marwondo, Marwondo&Setiana, Elia. (2020). Data Cleansing Strategies on Data Sets Become Data Science. International Journal of Quantitative Research and Modeling. 1. 145-156. 10.46336/ijqrm.v1i3.71.

[4] Huang, Fang. (2019). Data Cleansing. 10.1007/978-3-319-32001-4_300-1. In book: Encyclopedia of Big Data

[5] Geuder, N, and Quaschning, V. Soiling of irradiation sensors and methods for soiling correction. United Kingdom: N. p., 2006. Web. doi:10.1016/J.SOLENER.2006.06.001.

https://www.sciencedirect.com/science/article/abs/pii/S0038092X06001526?via%3Dihub

[6] Marko Schwandt, JuleRadtke, André Petersen, Anna Heimsath, and Richard Meyer , "Soiling impact on direct normal irradiation measurements", AIP Conference Proceedings 2126, 190014 (2019) https://doi.org/10.1063/1.5117711

[7] DeFreitas, Zoe, Ramirez, Anthony, Huang, Brian, Kurtz, Sarah, Muller, Matthew, and Deceglie, Michael G. Evaluating the Accuracy of Various Irradiance Models in Detecting Soiling of Irradiance Sensors: Preprint. United States: N. p., 2019. Web.

[8] Bacher, Peder& Madsen, Henrik &Perers, Bengt & Nielsen, Henrik. (2013). A non-parametric method for correction of global radiation observations. Solar Energy. 88. 13-22. 10.1016/j.solener.2012.10.024.

[9] Arima, Yusuke &Ooka, Ryozo&Kikumoto, Hideki. (2016). Bias correction method for solar radiation based on quantile mapping to provide weather data for building energy simulations. Journal of Environmental Engineering (Transactions of AIJ). 81. 1047-1054. 10.3130/aije.81.1047.

[10] Soulayman, Soulayman&Daude, N. &Darveesh, M. (1995). A Simple Method for Correcting Solar Radiation Measurements Made Using Non Calibrated Eppley-Type and Robitzsch-Type Pyranometers. International Journal of Solar Energy. 18. 125-135.

[11] M. Muller, L. Micheli and A. A. Martinez-Morales, "A Method to Extract Soiling Loss Data from Soiling Stations with Imperfect Cleaning Schedules," 2017 IEEE 44th Photovoltaic Specialist Conference (PVSC), 2017, pp. 2881-2886, doi: 10.1109/PVSC.2017.8366214.