www.rericjournal.ait.ac.th

Stochastic Model for Generating Synthetic Hourly Global Horizontal Solar Radiation Data Sets Based on Auto Regression Characterization

M.P. Anand*, [1], Athula Rajapakse*, and Bagen Bagen*, [+]

Abstract – *A large number of non-repetitive multi-year hourly solar radiation time-series dataset are desired when applying Monte Carlo techniques for planning and design of solar energy systems. Solar radiation models which utilize the clearness index and average value decomposition methods are commonly used to generate synthetic set of solar irradiances for this purpose. In this paper, a novel stochastic solar radiation model based on probability distributions of the first-order differences of hourly global solar horizontal radiation is proposed. The first-order differences are modeled using a trend component and a stochastic component represented using the cumulative distribution functions, both extracted from historical data taken over a window of 31 days around the considered day of the year. Measured solar radiation data from four different locations with varying climate characteristics were used to evaluate the proposed model in comparison to two previously reported models. The proposed method performed consistently better in terms of the similarity of probability distributions and autocorrelation functions, for all four locations and datasets.*

Keywords – *first order differences of solar radiation, long-term solar radiation models, Monte Carlo simulation, solar irradiance, synthetic hourly solar radiation.*

1. INTRODUCTION

The typical approach used for designing solar energy systems is to use a short period of historical weather data or Typical Meteorological Year (TMY) data for the particular location to evaluate the performance of the system being designed. However, a short period of historical data or TMY data may not capture the full stochastic nature of the variations in solar radiation over the long design life of solar energy systems. In order to overcome this limitation and reliably evaluate the feasibility of design alternatives and impact of different operational strategies associated with solar energy systems, probabilistic methods like Monte Carlo simulation (MCS) is used [1]. MCS, which involves a large number of trials, is a popular approach for evaluating systems involving random processes. Each trial simulation requires a distinct time series of solar radiation over the design life cycle of the system as an input to fully characterize the randomness. MCS typically requires hundreds, if not thousands, of trails before convergence, and the simulation input demands non-repetitive datasets of solar irradiance for each simulation trial [2]. Therefore, synthetic solar irradiance generators are often used to support MCS and similar probabilistic analysis approaches to perform design studies associated with siting and sizing of solar collectors, energy storage, etc. of both solar thermal and photovoltaic (PV) systems [3].

Jordan and Liu proposed a model to determine the horizontal surface diffuse radiation, the long-term average hourly and daily sums of diffused solar radiation considering varying levels of cloudiness [4]. An approach is proposed to generate sequences of global horizontal irradiance utilizing the probability of sunshine and cloudiness in [5]. Procedures for generating random but coherent sequences of daily global insolation values, and then decompose daily irradiance to hourly irradiance values are presented in [6], [7]. These models use a library of Markov transition matrices generated from long-term measurements for this purpose [8]. Similarly, a stochastic procedure has been proposed for generating synthetic sets of hourly solar radiation values in [9]. All these models primarily use the clearness index as a measure to capture the cloudiness during the sunshine hours while generating synthetic sets of solar radiation. Clearness index is calculated as the ratio of measured solar irradiance to the extraterrestrial solar radiation. The extraterrestrial solar radiation is calculated based on the solar geometry and is assumed to be having higher values compared to the measured solar radiation constituting the maximum value of k as 1 [7]. However, it has been reported in [10] that cloud reflected light added to the direct solar radiation could sometimes make the measured solar radiation values higher than the extraterrestrial solar radiation, especially near the sunrise and sunset. These instances can make the clearness index values greater than 1 and introduce errors to the models which use clearness index as a primary measure [10].

Reference [5] proposed to generate sequences of global horizontal irradiance utilizing the probability of sunshine and cloudiness models. According to the results, these models can reproduce the main characteristics of the natural time series of measured solar data. Many publications including [9], [11] further validate these models for different locations, and they

* Electrical and Computer Engineering, University of Manitoba, Winnipeg, Manitoba, Canada.

[+] System Planning, Manitoba Hydro, Winnipeg, Manitoba, Canada.

[1]Corresponding author:
Tel: +1-204-963-8765.
Email: mpa@myumanitoba.ca

generally tend to preserve the statistical characteristics in the solar radiation estimations. To generate synthetic sequences of daily and hourly global horizontal irradiation (GHI) a model is proposed in [12]. The model utilized synthetic daily and hourly models separately and reconcile them to match afterwards to generate synthetic solar radiation values. Additionally, some of the existing models intend to produce time series of hourly clearness index from the average values using stochastic procedures like average value decomposition [6] and Markov models [13], [14] to generate hourly clearness index from daily clearness index. However, solar system studies based on average values sometimes yield incorrect results [15].

There are attempts to develop solar irradiance generators capable of generating synthetic irradiance values at sub-hour time intervals for short periods such as 10-min [16], 5-min [17], 1-min [18] from mean hourly meteorological observation inputs. The motivation for developing these models are derived from the need of models that are essential in capturing the short-term temporal dynamics associated with solar radiation which can bring huge operational benefits such as real-time power system scheduling [19]. However, short-term synthetic solar irradiance generators are less applicable to long-term planning studies. Moreover, sizing, design, and performance evaluation of solar projects at any location under consideration require accurate and detailed information on both the stationarity and sequential properties of hourly solar radiation [20]. Analysis of the sequential characteristics of assorted meteorological variables, such as cloud cover, air temperature, wind speed, and relative humidity [15], [16], reveal very strong correlations at the one-hour time lag. Similar trend is revealed in a time series analysis of hourly GHI for a wide range of climatic stations that span temperate and tropic conditions [24]. The examination of the sequential behavior of hourly solar radiation revealed strong correlations over long persistence times nearly equal to the entire daylight period. The solar radiation time series input should reproduce not only the frequency of occurrence of various solar radiation levels for each hour of the day but also the persistence times and persistence strengths of hour-to-hour variations of radiation [20]. This is highly important with respect to the design of system including energy storage.

A simple model capable of generating synthetic, non-repetitive, long-term solar irradiation datasets that replicates both the stationary and sequential characteristics of measured radiation is helpful for long-term planning studies, especially for those using MCS based approaches. Since the solar radiation databases are becoming available for most locations in the world, such synthetic solar radiation data models should be easily developed using the solar radiation data for the interested location, measured over a sufficient period of time. Thus, the main goal of this paper is to develop a synthetic global horizontal solar radiation generation model that is usable for Monte Carlo simulation-based long-term energy planning studies. Followings are the specific objectives associated with this work:

1. To analyze the statistical properties of the difference in global horizontal irradiation measurements between two consecutive time steps, which will be referred to as the first order differences in GHI.
2. To develop a novel simple stochastic model for generating synthetic GHI data using the probability distribution functions of the first order differences in GHI.
3. To compare the proposed model with two previously published models [7], [9] and examine the ability of generating time-series of GHI data that can reproduce the statistical properties of the first order difference in GHI.

The rest of the paper is organized as follows: Section 2 provides the details regarding the measured solar data and primary locations used to build and test the proposed solar model. Section 3 explains the basic concept behind the first-order difference used to build the model, and Section 4 presents the proposed solar model, procedural flowchart used in generating the synthetic solar radiation using the proposed model. Section 5 describes two selected models from the existing literature for validating and comparing the results obtained using the proposed approach. Finally, simulation results have been demonstrated in Section 6 and Section 7 concludes the paper.

2. SOLAR RADIATION DATA

The measured hourly GHI data collected from four different locations with varying climate characteristics are used to develop and test the proposed model. Solar data from four different locations were collected, which includes two locations from Canada and two locations from the US. The Natural Resources Canada (NRC) [21], an open database is used to obtain the solar data of Canadian locations and Solar anywhere [22] is used to get the solar data of US locations.

The NRC's dataset [21] includes solar radiation data of 492 Canadian locations and solar values are primarily based on satellite-derived solar estimates using methods developed at State University of New York. As per the database information, the average standard error for all sites are calculated as 5%. Solar anywhere database [22] generates the solar radiation data from geostationary satellite images using the Perez Model [23]. The databased utilize satellite images to generate GHI measurements in resolutions as high as 1 km. As per the database, the standard error for annual solar radiation at all sites are close to 5%.

The distinction between the climate characteristics of the chosen locations are defined based on Köppen-Geiger climate classification [24], and these locations can be classified as subarctic, warm summer humid continental, temperate Mediterranean climate, warm oceanic climates respectively. The motivation behind choosing different locations is to ensure the unbiased (towards one particular dataset) and the generic nature of the developed model. Table 1 shows the locations

considered along with the database details and climate classification of the considered locations.

3. FIRST-ORDER DIFFERENCE IN GLOBAL HORIZONTAL SOLAR RADIATION

A strong correlation between the hourly solar radiation over long persistence time periods has been reported by [20]. Similarly, the autocorrelation values shown in Figure 1 (a) reveals the correlative structure of hourly GHI solar data measured at La Grange, USA. The strong correlation between the current hour solar radiation with those of three past hours can be seen in the partial correlation shown in Figure 1 (b). The relationship of this natural time series with respect to its immediate time steps are therefore evident from these correlation values. A similar trend in the autocorrelation/partial correlation values were observed for other locations considered in this work.

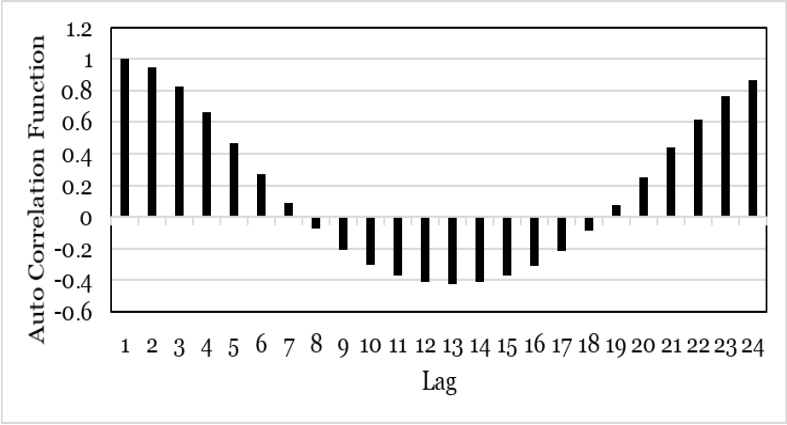
Therefore, a new property named as "First order difference in solar radiation" which simply represents the rise/fall of solar radiation at the current time step (t) with respect to the previous time step (t-1) is proposed in this work. The change in this global horizontal solar radiation between two adjacent time steps can be represented as:

$$\Delta H_G(t) = H_G(t) - H_G(t-1) \quad (1)$$

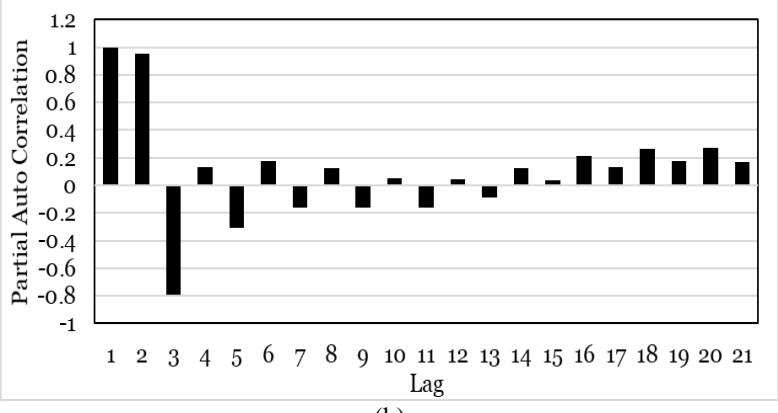
where, $H_G(t)$ is the global horizontal solar radiation measured at the current time step t and $H_G(t-1)$ is the value measured at the previous time step.

Table 1. Location details, available data period, sources of the database, and climate classification for the four locations considered in this work.

Location	Data period	Database	Köppen-Geiger climate classification
Division 23, Manitoba (DM-23) (Lat: 57.95, Long: -100.05)	January 2004 to December 2014	Natural resources Canada [21]	Subarctic
Division 21, Manitoba (DM-21) (Lat: 54.45, Long: -100.05)	January 2004 to December 2014	Natural resources Canada [21]	Warm summer humid continental
Leavenworth, Washington (LW-W) (Lat: 47.75, Long: -120.65)	January 2000 to December 2014	Solar anywhere [22]	Temperate Mediterranean climate
La Grange, Georgia (LG-G) (Lat: 41.65, Long: -104.45)	January 2000 to December 2014	Solar anywhere [22]	Warm oceanic climates



(a)



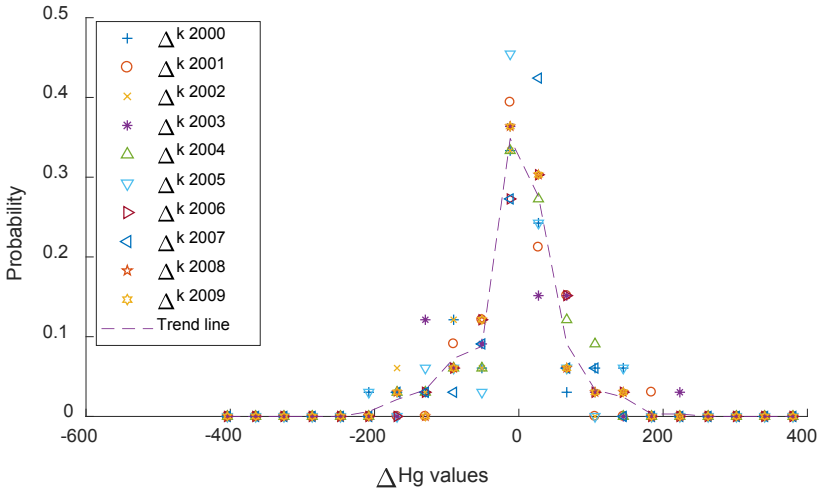
(b)

Fig. 1. (a) Autocorrelation and (b) partial correlation of hourly measured GHI at La Grange, USA.

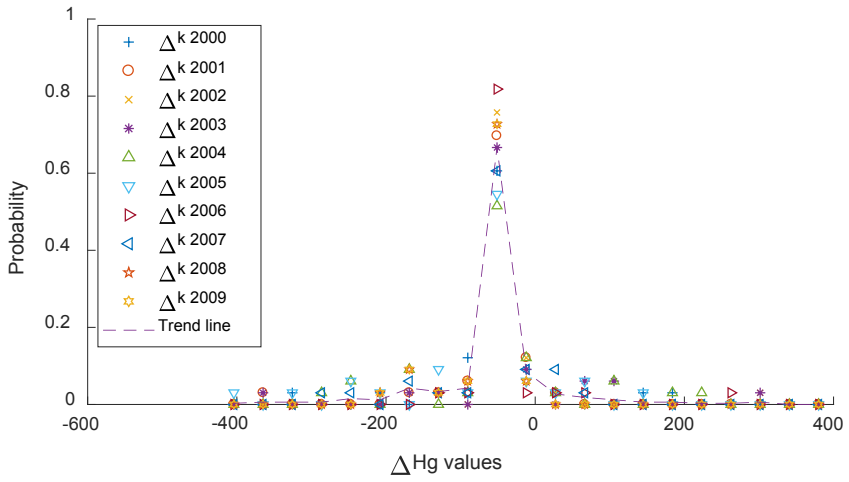
3.1 Annual ΔH_G Probability Distributions

In order to demonstrate the relevance of hourly ΔH_G values, the probability distributions of ΔH_G values calculated at three different time periods for 10 different years measured at La Grange, USA are shown in Figure 2. The trend line in Figure 2 represents the distribution when all 10 years of data are considered together. Figure 2 shows that ΔH_G values at a given hour in a given date has a characteristic distribution that does not change significantly from year to year. Similar distributions could be observed for ΔH_G values computed for other time periods too. Almost identical probability values for different years point to the fact that even though random variation exists with respect to measured solar radiation, annual probability distributions of ΔH_G values are similar and predictable. Even though the primary location considered in Figure 2 is La Grange, USA, a similar pattern of ΔH_G probability values were observed for other locations such as Leavenworth, in the USA and Division 23/21 in Manitoba, Canada.

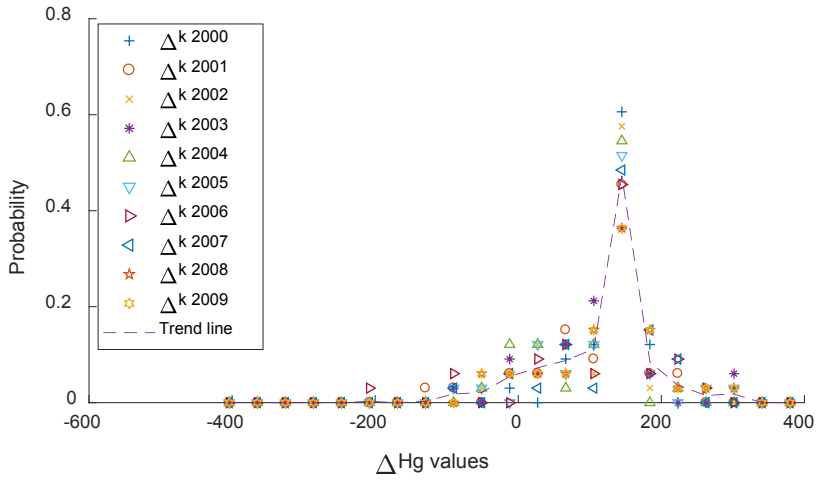
To quantify the similarity of the distributions of ΔH_G values, we used two-sample Kolmogorov Smirnov test (KS-II) with varying confidence levels. The test was applied to all possible distinct pairs of years (10 years of data results in 45 distinct pairs) taking ΔH_G values for the respective pair of years as the samples from the two distributions. Figure 3 shows the passing rates of KS-II test for the four different times considered in Figure 2. Over 95% pairs of years passed the test (*i.e.* confirmed the null hypothesis that two distributions are the same) even when a 95% confidence level is considered. Therefore, it can be deduced that this characteristic distribution of ΔH_G values can be a useful metric to compare the accuracy of prediction models while comparing with the measured solar radiation data. In other words, any good synthetic solar irradiance generator should be able to preserve these characteristic distributions, in addition to the other statistical properties commonly considered.



(a)



(b)



(c)

Fig. 2. Comparison of inter-annual ΔH_G probability distributions for ten different years for periods (a) 14:00-15:00 (b) 11:00-12:00 (c) 15.00-16.00 measured at La Grange, USA.

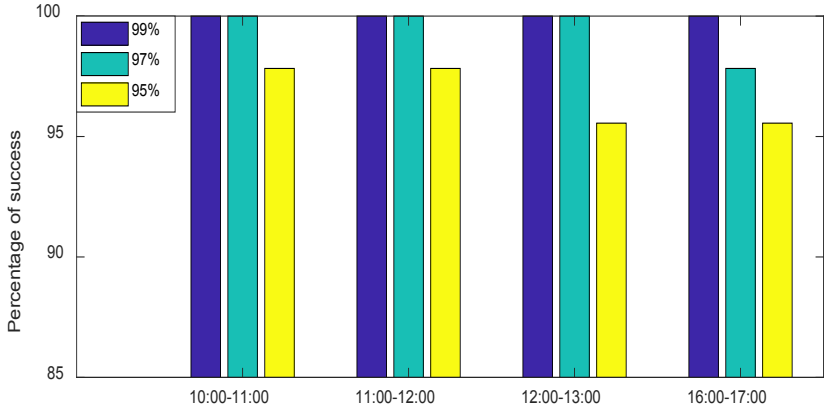


Fig. 3. Percentage of pairs of years that passed the KS-II test for asserting the similarity between the distributions of annual ΔH_G values computed at four different time periods measured at La Grange, USA.

3.2 Inter-Annual ΔH_G Probability Distributions

To examine the feasibility of utilizing the "first-order differences" in generating synthetic sets of solar radiation, the intra-annual characteristics distribution of the first-order differences were obtained considering ΔH_G values from a 31-day window around the given date. These distributions for 10 different years are shown in Figure 4. It can be observed that the probability distributions for different years are similar, although the random deviations among them are higher when considered a 31-day window, instead of the whole year. Since these distributions are generated using a fewer number of samples of ΔH_G values such differences are expected. A similar pattern of ΔH_G values were observed while considered other locations such as Leavenworth, in the USA and Division 23/21 in Manitoba, Canada.

Similar to the annual ΔH_G probability distributions, KS-II test was used to quantify the similarity of the probability distributions with varying confidence levels as shown in Figure 5. It is evident from the figure that over 90% passing rate was observed for KS-II test, for all pairs of years, at all three confidence levels considered. Therefore, it is proposed to use a window of 31 days centered around the current day to obtain the probability distributions of the first-order differences ΔH_G values) for developing a synthetic hourly solar radiation data generation model. The characteristic obtained with a 31-day window would be more consistent with monthly average characteristics that are typically considered in most solar radiation models [7], however, other window lengths are possible. Experiments with 15, 21, and 41 day window lengths showed that adequately smooth probability distributions can be obtained. However, shorter the window length, smaller number of data points in the probability distribution functions.

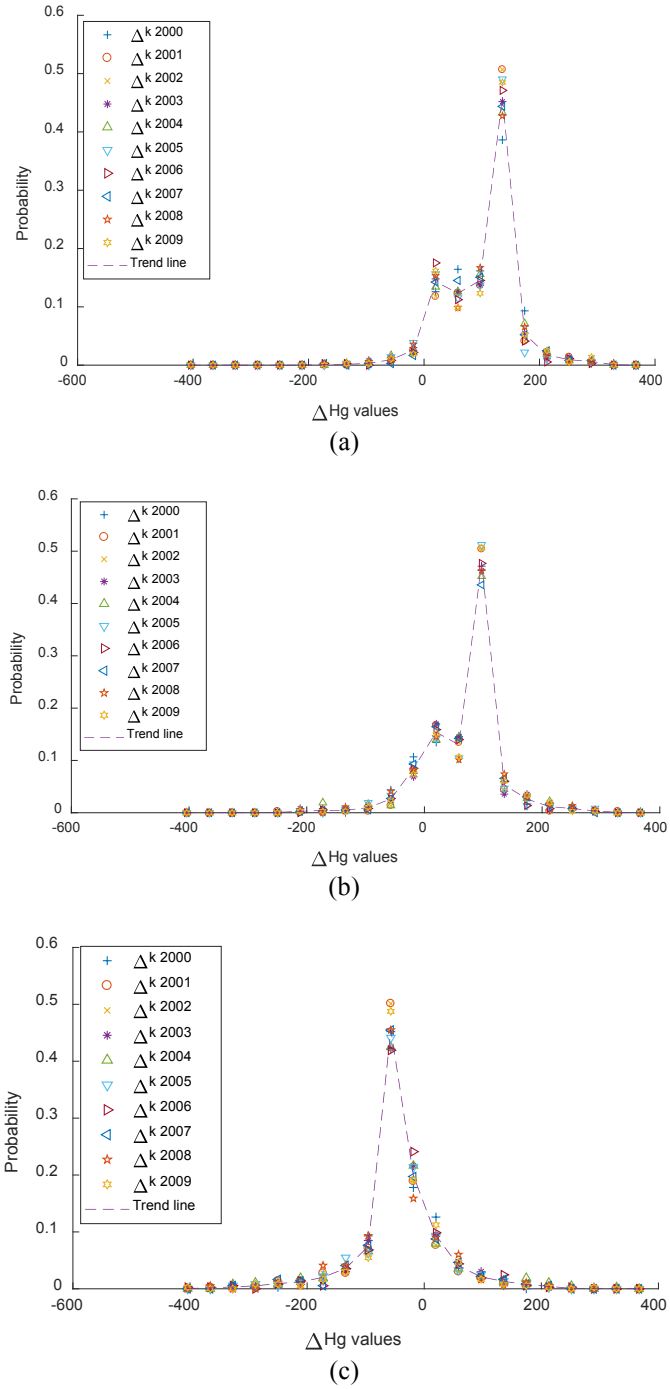


Fig. 4. Comparison of inter-annual ΔH_G probability distributions for ten different years for the specific day and period (a) Feb 12, 14:00-15:00 (b) May 14, 11:00-12:00, and (c) Dec 15, 14.00-15.00, measured at La Grange, USA.

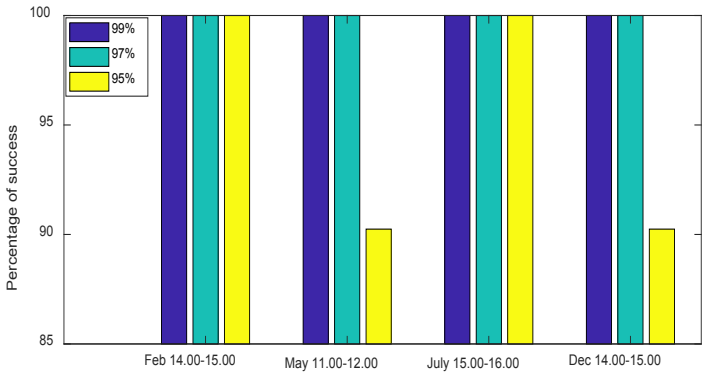


Fig. 5. Percentage of pairs of years that passed the KS-II test for asserting the similarity between the distributions of intra-annual ΔH_G values computed at four different time periods.

3.3 Modeling ΔH_G using the Trend and Stochastic Components

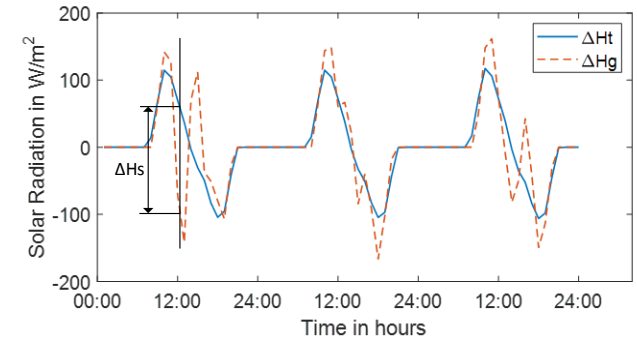
The first order differences ΔH_G can be considered as a combination of trend and stochastic component as shown in (2). A similar approach has been used by [25], [26] in the past.

$$\Delta H_G(t) = \Delta H_T(t) + \Delta H_S(t) \quad (2)$$

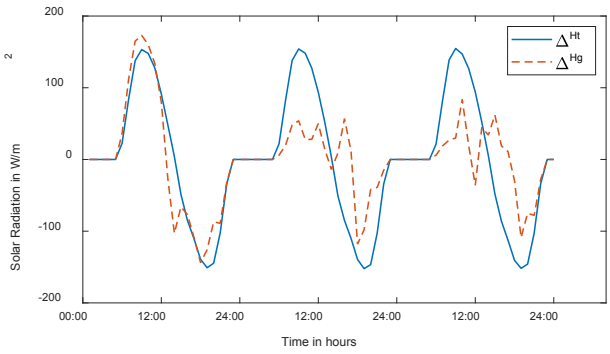
The trend component ΔH_T represents the general trend in the pattern of first-order differences in solar radiation and the stochastic component ΔH_S represents the random deviations from the trend component. The trend component $\Delta H_T(t)$ for hour t in a given day can be obtained from the historical data. It is the mean of $\Delta H_G(t)$ values observed during a 31-day sliding window centered on the given day typically averaged over several years. Once the trend component is determined, the stochastic component $\Delta H_S(t)$ can be calculated from historical data:

$$\Delta H_S(t) = \Delta H_G(t) - \Delta H_T(t) \quad (3)$$

Figure 6 shows some sample time series plots of ΔH_G, ΔH_T and ΔH_S values for different days in a year. The trend component ΔH_T values are calculated beforehand based on the measured solar radiation and CDFs are used to model the stochastic component ΔH_S determined using (3).



(a)



(b)

Fig. 6. Sample time series plot illustrating the trend (ΔH_T) and actual (ΔH_G) component of first order difference in solar radiation for (a) March 12-14, and (b) July 23-25.

4. FIRST-ORDER DIFFERENCE IN GLOBAL HORIZONTAL SOLAR RADIATION

A good synthetic solar radiation generation model should be able to:

- Follow the characteristic diurnal and seasonal variations,
- Include the random variations due to clouds and atmospheric conditions,
- Preserve statistical measures such as mean, standard deviation, etc.,
- Preserve correlation among consecutive time steps.

In contrast to most previously published approaches that primarily model the solar radiation expected at a given hour, the new model proposed in this paper attempts to model the expected change in solar radiation at a given hour, ΔH_G. Then the solar radiation at hour t, $H_G(t)$ is obtained by adding the change of solar radiation generated for hour t, $\Delta H_G(t)$ to the previous hour solar radiation $H_G(t-1)$ as shown in (4). By continuing this process, a time series of solar radiation data can be obtained in a recursive manner.

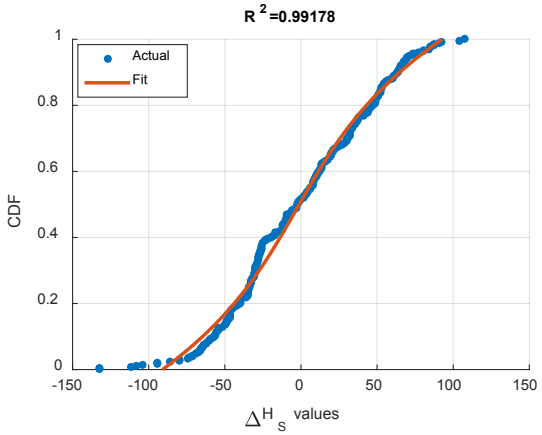
$$H_G(t) = H_G(t-1) + \Delta H_G(t) \quad (4)$$

The proposed model is developed based on the following hypotheses:

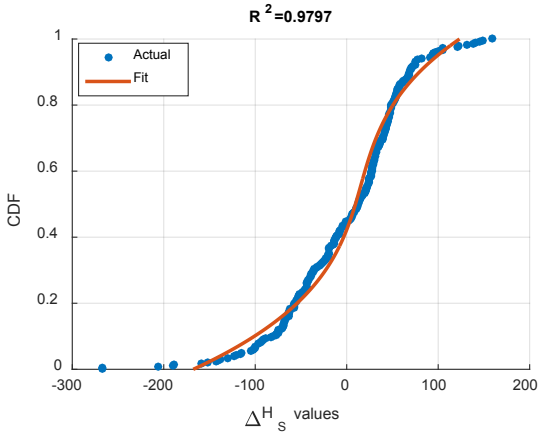
- Due to the sinusoidal like variation of clear sky radiation, solar radiation values of two consecutive time steps are highly correlated. Thus, the current hour solar radiation is highly relevant when generating the next hour solar radiation and forms the basis for (4).

- Characteristics of solar radiation within a 31-day period centered around the current day are similar. Thus, cumulative distribution functions required for the model are developed on a daily basis with a sliding window of 31 days similar to [12].
- Annual first-order differences of the hourly solar radiation representing the trend and random component, calculated at a given hour, for different years, has a unique distribution that does not change from year to year. Therefore, synthetic values generated from a good solar model would preserve this characteristic distribution.

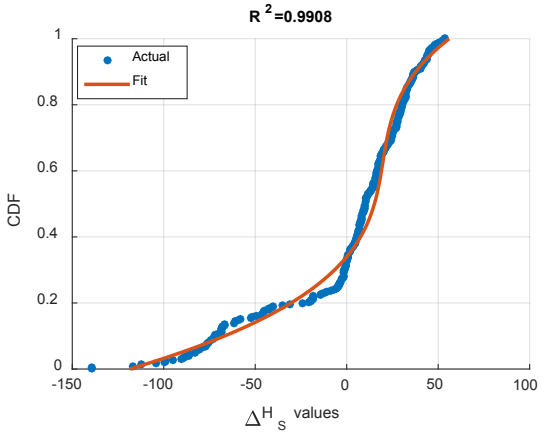
As per the hypothesis derived, trend and random components of solar radiation are needed to determine the hourly synthetic solar radiation at any time period t. The trend component ΔH_T is an averaged value over the years with a 31-days moving window while random component ΔH_S are modeled and generated using probability distribution functions (CDFs). Several examples of CDFs of ΔH_S are shown in Figure 7.



(a)



(b)



(c)

Fig. 7. Cumulative distribution functions (CDFs) of ΔH_S values developed for (a) Jan 25, 11.00-12.00, (b) March 15, 16.00-17.00, and (c) September 10, 18.00-19.00 measured at La Grange, USA.

4.1 Building of the Synthetic Solar Radiation Model

Since the model need to store a large number of CDFs to generate the ΔH_S values, it is more computationally efficient to fit CDF data to a model and store the model coefficients. It was found that polynomial functions are appropriate to approximate the inverse relationships of the CDFs. Trials with the data for the selected locations showed that 3rd order polynomials, as given in (5), result in Pearson correlation coefficients greater than 0.95 for all CDFs.

$$\Delta H_S(x) = a_0^t + a_1^t x + a_2^t x^2 + a_3^t x^3 \quad (5)$$

where $x \in [0-1]$ is the cumulative probability, and $\Delta H_S \in [\Delta H_{S,min}(t), \Delta H_{S,max,}(t)]$ at hour t. The coefficients $a_0^t, \ldots, a_3^t$ are found by least-square fitting [27]. Other approaches such as bootstrapping the error of data used for obtaining CDFs can be applied to generate random samples of ΔH_S values. However, if measured data for a sufficiently long period of time (>10 years) is available, with a 31 days window, there is enough data samples to generate a reliable CDF. Therefore, fitting a curve to represent a CDF is a more computationally efficient approach. Figure 8 demonstrates the step-by-step procedure followed in building the model. The following steps explain in detail the procedure carried out in building the proposed model using the measured solar radiation.

Step-1.1: From the historical data, calculate ΔH_G values, and then considering a 31-day sliding window, calculate the trend component (ΔH_T) for each hour of the year. Using the trend component, calculate the stochastic component (ΔH_S) values.

Step-1.2: Segregate ΔH_S values corresponding to a given hour t (within the 31-day sliding data window, of all years) into bins. Develop the CDFs $F(\Delta H_S)$ for each hour.

Step-1.3: The coefficient values of the 3rd order polynomials for each hour t is calculated using the least square estimation.

After constructing the model, the estimated coefficient values are stored in an 8760×4 array. Note that all coefficients corresponding to nighttime hours are zero, indicating a 100% probability of having $\Delta H_S = 0$ in these hours. However, these zero coefficients are also kept in the array for the convenience of programming.

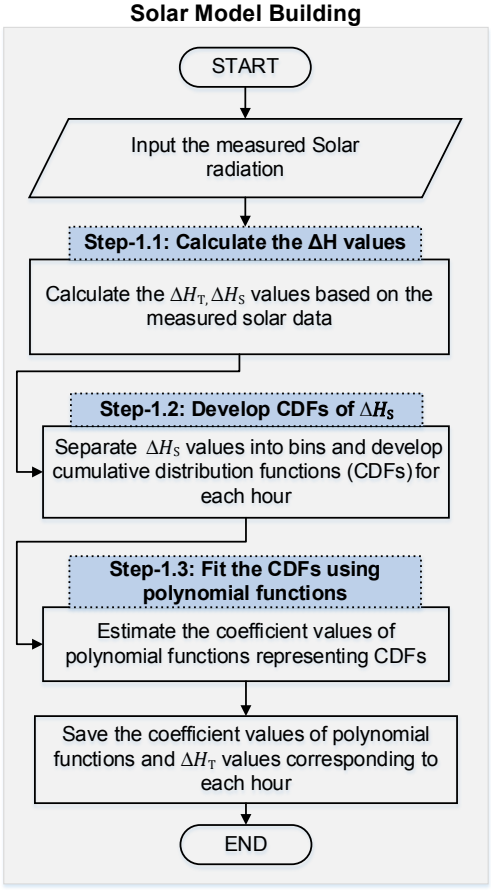


Fig. 8. Procedural flowchart for building the proposed synthetic solar radiation model.

4. 2 Generation of Synthetic Solar Radiation Data

To generate synthetic hourly solar radiation data, the coefficient values, synthetic data starting date/time and end date/time are given as inputs. Figure 9 illustrates the flowchart used in generating synthetic hourly global horizontal solar radiation. The following are the simple steps followed to complete the data generation.

Step-2.1: Recall the CDF coefficients (a_i^t) corresponding to the current hour (t). Then draw a uniformly distributed random number in the range [0.0 - 1.0] and use it as the cumulative probability in (5) to obtain a value for the stochastic component of the first-order difference in solar radiation (say $\Delta\widehat{H_S}(t)$). The initial value of $H_G(t-1)$ in (4) is assumed be to zero due to the sunset hour. Therefore, the initial value of solar radiation $H_G(t)$ also will be zero.

Step-2.2: The estimated $\Delta\widehat{H_S}(t)$ value is added to the trend component of solar radiation $\Delta H_T(t)$ to estimate the first-order difference in solar radiation $\Delta\widehat{H_G}(t)$ as in (2).

Step-2.3: To estimate the solar radiation of hour t, the generated first-order difference in solar radiation $\Delta\widehat{H_G}(t)$ is added to the previous hour solar radiation $\widehat{H}(t-1)$ as in (4). If the estimated solar radiation $\widehat{H_G}(t)$ is out of the range $[H_{G,min}, H_{G,max}]$, the procedure is repeated by randomly drawing another $\Delta\widehat{H_S}(t)$ until $\Delta\widehat{H_G}(t)$ will become within the limits. $H_{G,min}$ and $H_{G,max}$ represents the minimum and maximum solar radiation for hour t observed in the historical data (31-day moving window). Continue the procedure until solar radiation generation has been completed for the required duration.

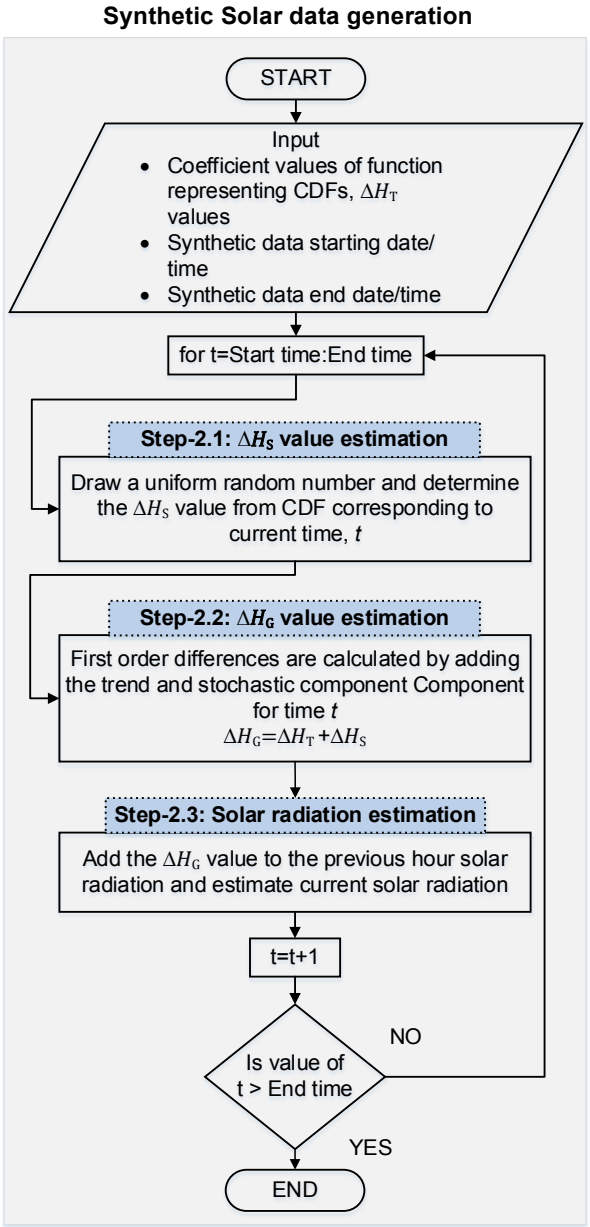


Fig. 9. Procedural flowchart for generating synthetic solar data.

5. RESULTS AND DISCUSSIONS

In this section, the simulated results are compared with measured solar radiation data and existing solar models in terms of comparison plots, autocorrelation, probability distributions, and mean values. All the simulated results were computed using the commercial software package MATLAB R2017a [28].

5.1 Description of Models used for Comparison

There are many stochastic solar models developed in the past that are capable of generating synthetic solar

dataset. Among many of those, the models proposed in [7], [9] account for the autocorrelation between hourly solar radiation in the modeling of solar radiation through a stochastic disaggregation procedure to convert daily solar radiation to hourly solar radiation. Additionally, the models have used clear sky indices as a primary factor to develop the model and average-decomposition procedures are utilized to estimate the hourly clearness index from daily and/or monthly clearness index [6], [7]. To compare the performance of models which primarily use k-values and decomposition procedures to build synthetic solar irradiance generators with the proposed approach, the authors found these models are representative. Moreover, these models have been validated using the solar radiation data collected from Europe, the USA and Canadian cities [8], [9]. The proposed model has also used the development/testing of the model utilizing the measured solar data collected from the USA and Canada. For the convenience of representation, these models are represented as SHSRG [9] and SISIM [7] models in the future sections. The results associated with these models are used to compare the efficiency of the proposed model.

5.2 *Training the Model and Generation of Synthetic Data*

Figure 10 illustrates a time series plot of a part (three years) of the long-term measured hourly solar radiation data used for building the model (plotted in blue) followed by simulated hourly solar radiation data using the developed model (plotted in orange). It can be observed from Figure 10 that the designed model is able to maintain the general trend of actual data over the time period of one year. Since the proposed model is not meant to predict the solar radiation at a given hour on a given day, it is highly important to compare the main descriptive parameters such as autocorrelation, probability distributions, and average values of the simulated and actual solar radiation [29]. Also, it is important to investigate whether the model captures diurnal and seasonal variation on average.

5.3 *Autocorrelation and Partial Autocorrelation of Hourly Solar Radiation Time Series*

There is a strong correlation between the consecutive hourly solar radiation values. The autocorrelation and partial correlation functions of the measured solar data shown in Figures 11 and 12 underlines this fact. Both SHSRG [9] and SISIM [7] solar radiation models consider the correlation between clearness indices of consecutive days and hours in generating the random solar radiation values. Additionally, the models try to preserve the time correlation by considering previous hour solar irradiance values while generating the current hour irradiance values. Based on this character, it is expected that these models can preserve the time correlation of natural time series of measured solar radiation. However, the generated solar radiation data from SHSRG and SISIM models shows a lower correlation compared to the correlation seen in the measured radiation at all lags for both locations shown in the plots.

On the other hand, the autocorrelation and partial correlation functions of the simulated data from the proposed model are much more similar to those of the measured solar data. This is because, in the proposed model, the current hour solar radiation is calculated as the previous time-step solar radiation $\widehat{H}(t-1)$ plus the difference $\Delta\widehat{H_G}(t)$. Hence, each hour solar radiation is calculated considering a sequence of past solar radiations values. Therefore, the model automatically incorporates the relationship of current hour solar radiation value to all past values in the day. This is inherent in the model and contribute to preserve the auto-correlation property for several time-lags. These comparisons underline the fact that the proposed solar radiation model better preserves the time correlation properties and performs better than SISIM/SHSRM models in this aspect.

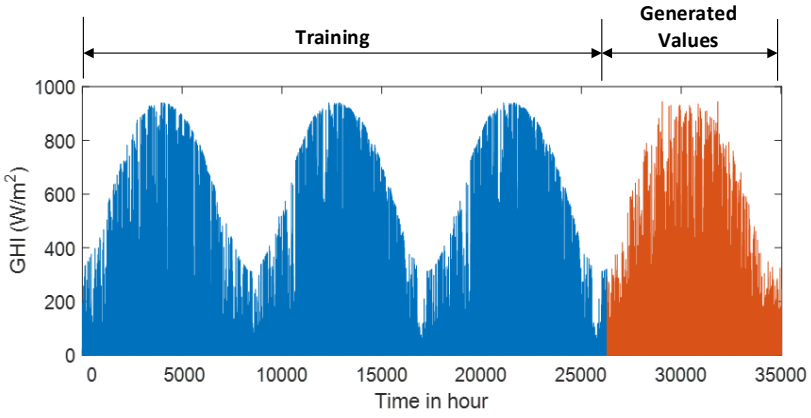
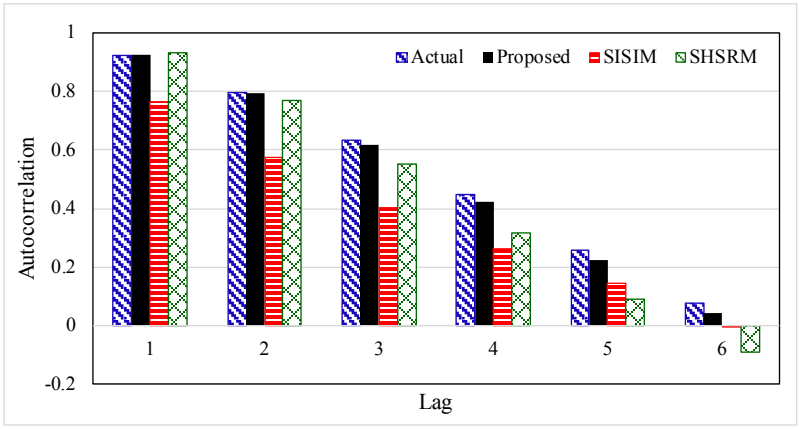
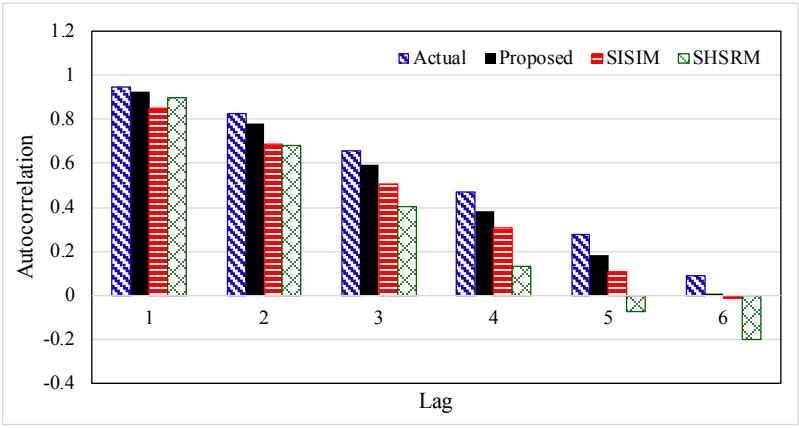


Fig. 10. Training and generated values of solar radiation using the proposed solar model for La Grange, USA.

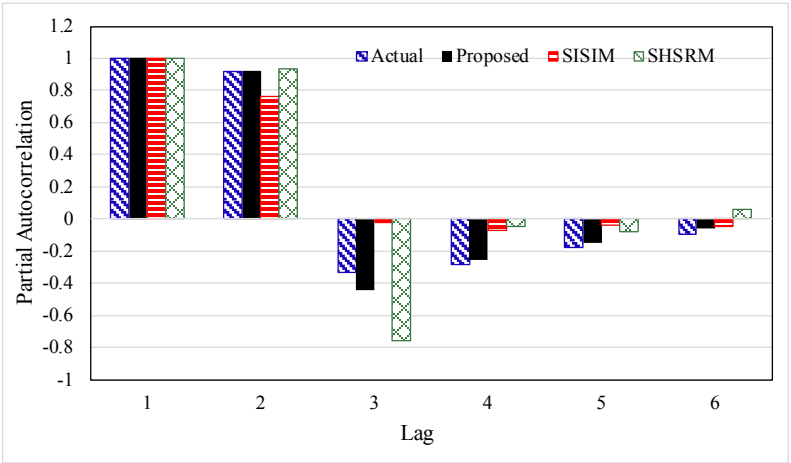


(a)

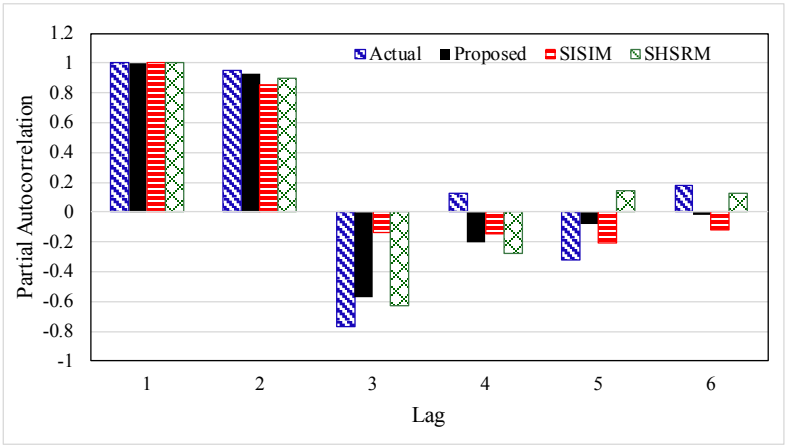


(b)

Fig. 11. Comparison of autocorrelation of measured hourly GHI values and generated hourly GHI values using the proposed model, SISIM, and SHSRM for (a) La Grange, USA and (b) Division 23, Manitoba, Canada.



(a)

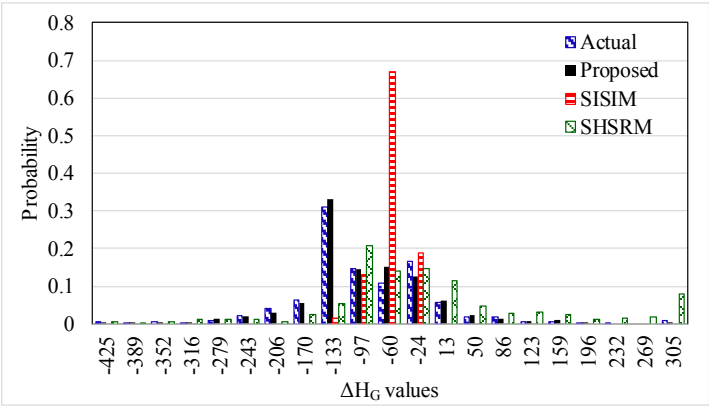


(b)

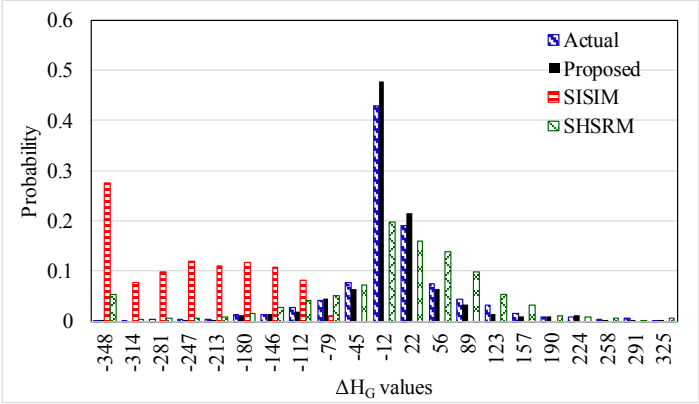
Fig. 12. Comparison of partial autocorrelation of measured hourly GHI values and generated hourly GHI values using the proposed model, SISIM, and SHSRM for (a) La Grange, USA and (b) Division 23, Manitoba, Canada.

5.4 Probability Distributions

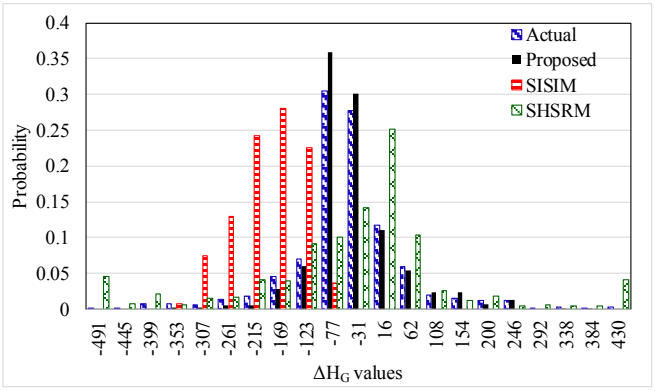
It is expected that the annual probability distributions of ΔH_G values computed from synthetic solar radiation values from a good model also demonstrate the same distribution that was found from the measured data illustrated in Figure 2. Figure 13 compares the distributions of ΔH_G values obtained from the solar radiation measurements with those obtained from the synthetic data generated using the proposed model, SISIM model, and SHSRM model, for four different locations at four different time periods. It can be seen from Figure 13 that the distributions of ΔH_G values obtained SISIM and SHSRM models significantly differ from the ΔH_G distributions of the actual solar radiation data. As expected, the probability distribution functions of ΔH_G are similar for the actual and synthetic data generated from the proposed model, because the proposed model for hourly global horizontal radiation was developed based on the probability distribution functions of the first-order differences of the historical hourly radiation measurements.



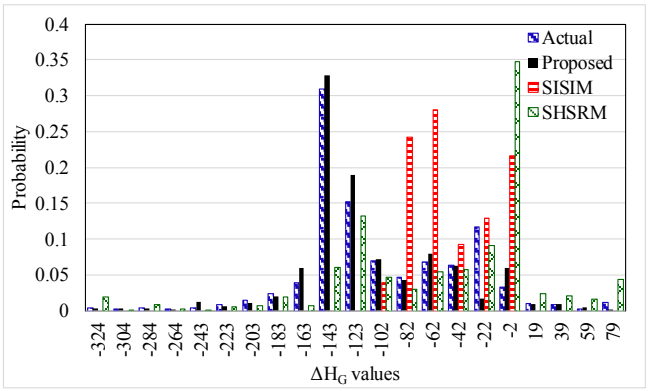
(a)



(b)



(c)



(d)

Fig. 13. Comparison of the probability distributions of ΔH_G obtained from the solar radiation models (SHSRM and SISIM) with those of actual data for the periods (a) 10.00-11.00 at Division 23, Manitoba, Canada, (b) 15.00-16.00 at Division 21, Manitoba, Canada, (c) 11.00-12.00 at Leavenworth, USA, and (d) 15.00-16.00 La Grange, USA.

The Euclidean distance between two ΔH_G probability distributions can be used to quantify the similarity of distributions. Euclidean distance between two distributions can be defined as shown in (6).

$$d(x,y) = \sqrt{\sum_{l=1}^{P}(x_l - y_l)^2} \quad (6)$$

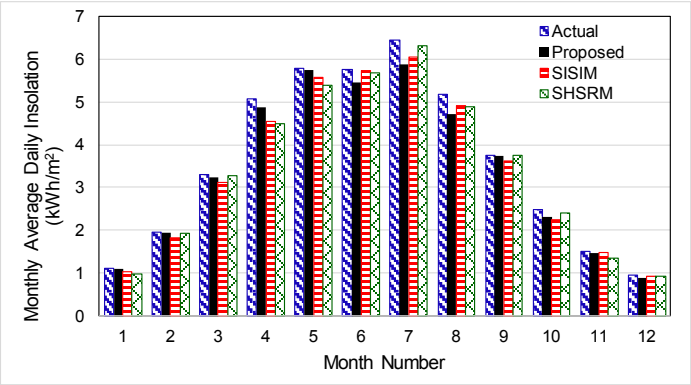
where, d represents the Euclidean distance (ED), x_l indicates the probability values corresponding to the actual ΔH_G distribution, y_l are probability values corresponding to the simulated ΔH_G distribution values, and P indicates the total number of ΔH_G points in the x-axis. The lower the ED, higher the similarity between two distributions.

Table 2 shows the Euclidean distances between the ΔH_G distributions obtained from the measured solar radiation and the synthetic data generated using the proposed, SISIM, and SHSRM models. Table 2 shows the similarity values of hourly probability distributions and the average similarity for all time periods, for four different locations. The highest average ED value is seen for SISIM model. On the other hand, the proposed model has the lowest average ED value for all four locations. This verifies the ability of the proposed model to produce ΔH_G distributions which are more similar to those corresponding to the actual measured solar radiation data.

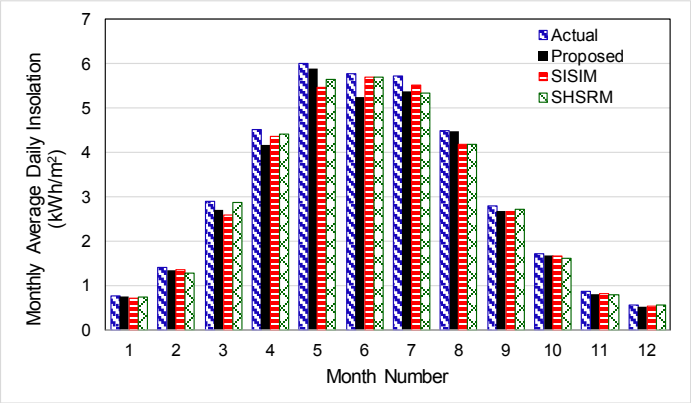
5.5 Monthly Mean Daily Insolation and Monthly Mean Hourly Solar Radiation

Comparisons of the simulated and actual solar radiation data obtained from the models described previously, and the proposed model is shown in Figures 14 and 15 for monthly mean daily solar energy calculated for four different locations. These averages are taken over 10 years. It can be observed that all the models fairly capture the monthly mean. Even though the proposed model is not characterized to explicitly capture the daily or hourly mean values like [7], [8], it is clear from the figure that simulated and measured hourly global horizontal radiation values are very close.

Additionally, the comparisons of monthly mean hourly solar radiation between measured solar radiation, proposed model, SISIM and SHSRM calculated is shown in Figures 16 and 17. The values are shown for four different locations and four different seasons where the average is taken over 10 years. It can be seen from the figure that all models were able to preserve the hourly averages close to the measured solar data.

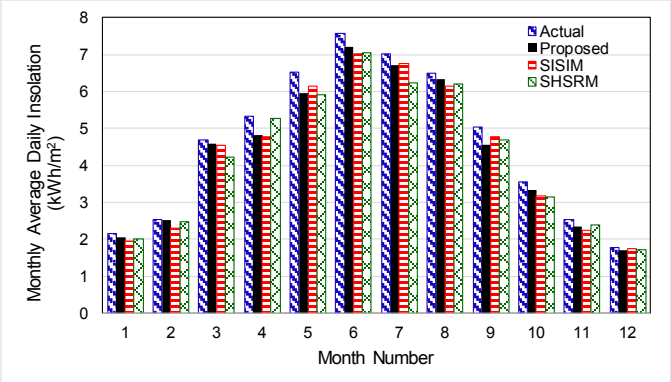


(a)

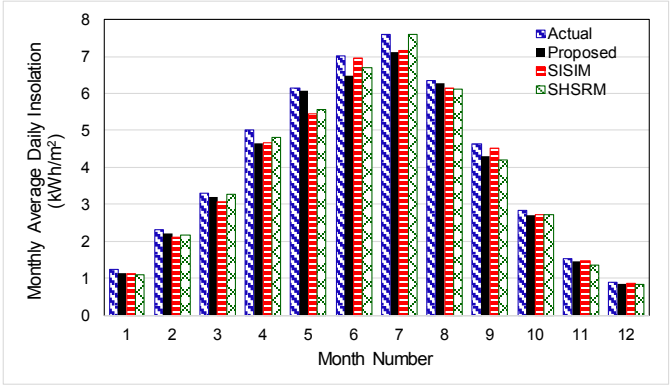


(b)

Fig. 14. Comparison of monthly mean daily insolation of measured solar radiation, the proposed, SISIM, and SHSRM models for (a) Leavenworth, USA and (b) La Grange, USA.



(a)

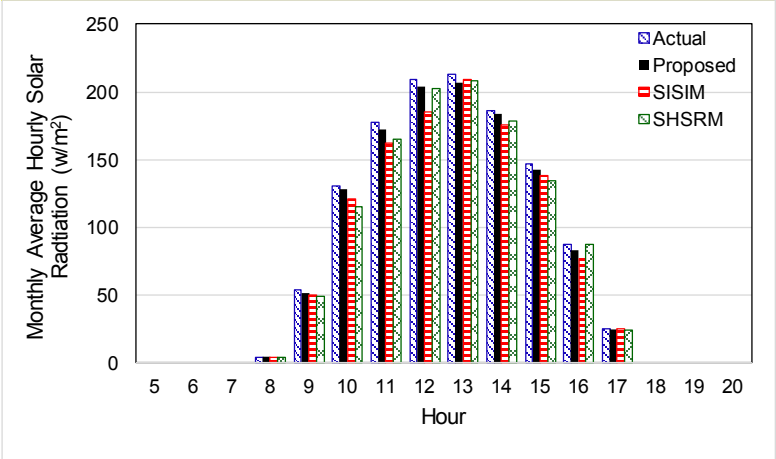


(b)

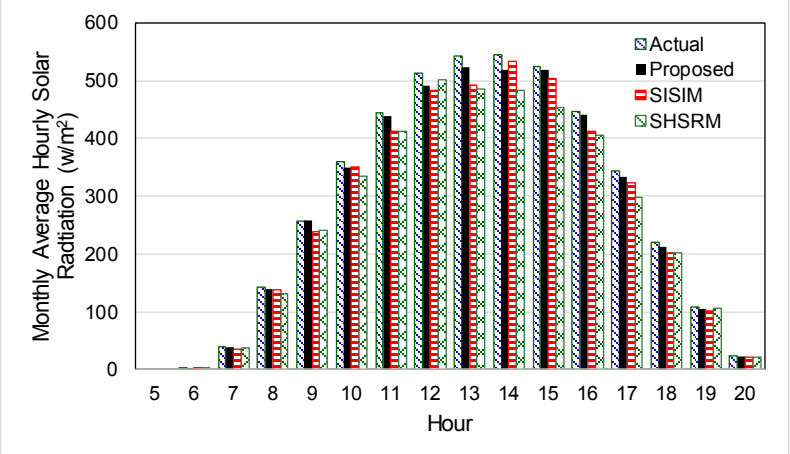
Fig. 15. Comparison of monthly mean daily insolation of measured solar radiation, the proposed, SISIM, and SHSRM models for (a) Leavenworth, USA and (b) La Grange, USA.

Table 2. Euclidean distances between the ΔH_G distributions of the actual data and the calculated data (using the proposed, SISIM, and SHSRM models) for different locations.

Time Period	DM-23			DM-21			LW-W			LG-G		
	Proposed	SISIM	SHSRM	Proposed	SISIM	SHSRM	Proposed	SISIM	SHSRM	Proposed	SISIM	SHSRM
8	0.10	0.19	0.16	0.44	0.57	0.58	0.09	0.39	0.40	0.44	0.57	0.58
9	0.06	0.14	0.16	0.11	0.49	0.41	0.07	0.66	0.40	0.11	0.49	0.41
10	0.06	0.64	0.28	0.06	0.36	0.26	0.05	0.35	0.30	0.07	0.37	0.27
11	0.07	0.78	0.30	0.08	0.55	0.19	0.10	0.57	0.28	0.08	0.56	0.22
12	0.06	0.35	0.33	0.05	0.62	0.26	0.06	0.57	0.40	0.06	0.62	0.27
13	0.10	0.29	0.11	0.05	0.77	0.49	0.09	0.42	0.20	0.06	0.77	0.49
14	0.06	0.61	0.07	0.06	0.28	0.42	0.09	0.74	0.29	0.08	0.29	0.42
15	0.11	0.43	0.25	0.10	0.30	0.35	0.06	0.90	0.30	0.12	0.32	0.35
16	0.19	0.32	0.86	0.13	0.32	0.31	0.13	0.32	0.20	0.13	0.31	0.31
17	0.16	0.53	0.61	0.15	0.32	0.33	0.07	0.57	0.18	0.15	0.31	0.33
18	0.11	0.31	0.37	0.15	0.24	0.58	0.06	0.32	0.18	0.15	0.36	0.58
19	0.11	0.57	0.31	0.09	0.48	0.13	0.09	0.46	0.17	0.07	0.27	0.13
20	0.10	0.40	0.16	0.15	0.31	0.15	0.04	0.22	0.24	0.13	0.30	0.15
21	0.11	0.20	0.06	0.00	0.00	0.33	0.01	0.04	0.22	0.00	0.00	0.33
22	0.00	0.00	0.13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Average of all time periods	0.09	0.38	0.28	0.11	0.37	0.32	0.07	0.43	0.25	0.11	0.37	0.32

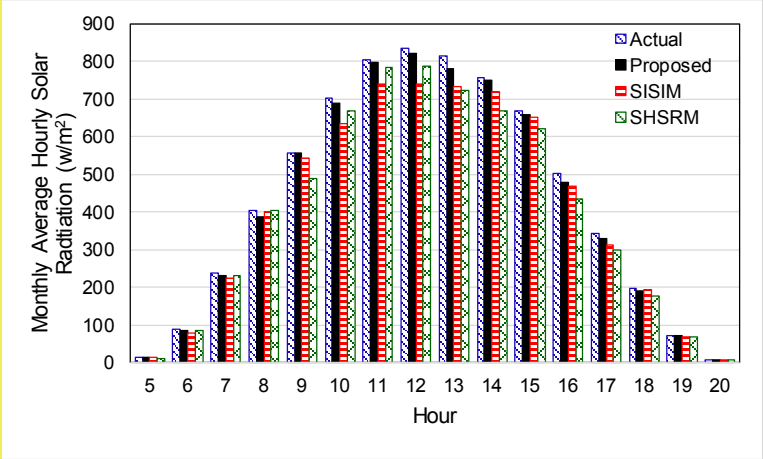


(a)

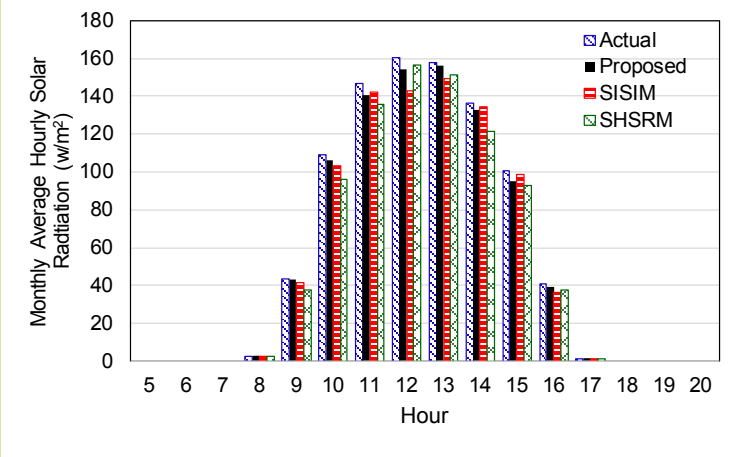


(b)

Fig. 16. Comparison of monthly mean hourly solar radiation of measured solar radiation, the proposed, SISIM, and SHSRM models for (a) Division 23, Manitoba, Canada and (b) Division 21, Manitoba, Canada.



(a)



(b)

Fig. 17. Comparison of monthly mean hourly solar radiation of measured solar radiation, the proposed, SISIM, and SHSRM models for (a) Leavenworth, USA and (b) La Grange, USA.

5.6 Testing the Generic Applicability of the Model

The new model is proposed for generating synthetic solar radiation data for long-range simulations for a given location. The model parameters are obtained from the available measured data for the location considered. Since there is no normalization procedure involved in the proposed model, it is not expected to be a generic model that can cover very different geographical areas. This is evident from the samples of the cumulative probability distribution functions of ΔHs values for different locations for the same time intervals shown in the Figure 18. Authors considered the feasibility of developing a solar radiation model considering the first-order differences of the hourly clearness index (Δk values), however, the natural variation for k-values over a day is obscured by the very high k values (occasionally exceeding 1.0) at the beginning and end of the daylight hours. The values of measured and extraterrestrial solar radiation are very close at very low solar altitude angles, and sometimes the measured global horizontal solar radiation is higher than the calculated extraterrestrial solar radiation due to cloud reflected radiation and possible measurement errors. This phenomenon usually contributes towards unusually high peaks in Δk variations at the beginning and the end of the daylight hours.

Moreover, the physical interpretation of Δk values is not as straightforward as ΔH values. Therefore, the authors preferred to develop the model based on the absolute differences in solar radiation. However, further research may enable solving these modeling challenges and development of more generalizable models based on the concept of first-order differences of the clearness index.

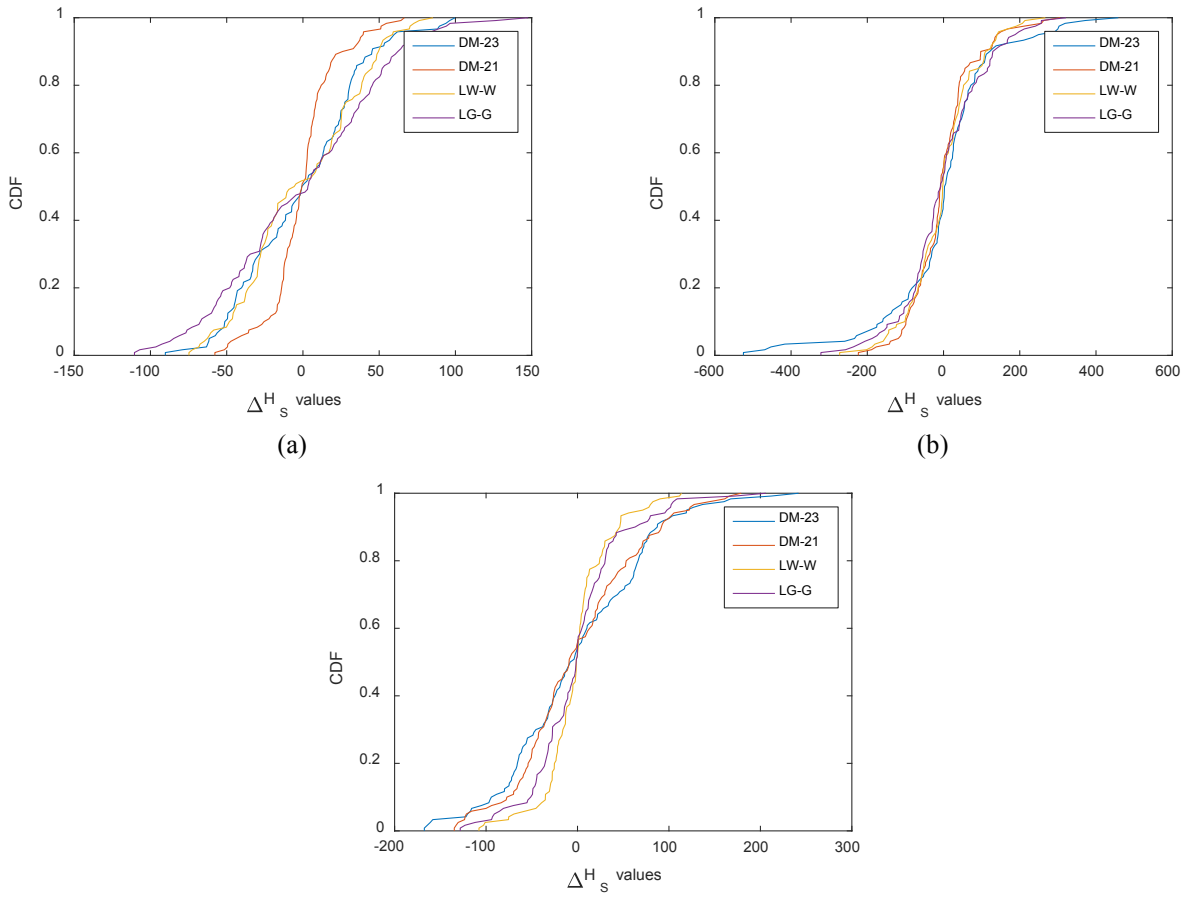


Fig. 18. Comparison of ΔHs CDFs developed based on the measured solar radiation from four different locations (a) Jan 30, 10.00-11.00, (b) April 23, 13.00-14.00 and (c) Sep 25, 10.00-11.00.

6. CONCLUSION

Analysis of historical hourly global horizontal solar radiation data revealed that the first-order differences of solar radiation calculated at a given hour has a characteristic distribution that does not tend to vary from year to year. A new long-term stochastic solar radiation generation model that utilizes this characteristic distribution function of the first-order differences of hourly solar radiation was proposed. It was shown that the first-order differences of solar radiation of a particular hour in year can be represented in terms of a trend component and a stochastic component characterized using a cumulative distribution function obtained from the historical data.

As per the analysis presented in the paper using data for four distinct locations, synthetic solar radiation generated SISIM and SHSRM solar radiation models failed to accurately reproduce the distribution functions for the first-order differences of global horizontal solar radiation. In contrast, the proposed model closely reproduced the probability distributions of the first-order differences, while outperforming the existing models. Although the generalizability of the developed model to different locations is limited, building the model from historical data for a given location is simple and straightforward, and useful for applications such as Monte Carlo simulations.

ACKNOWLEDGEMENT

The authors would like to thank Manitoba Hydro, Manitoba, Canada and Mitacs Accelerate Program, Canada for the financial support provided to successfully complete the research.

REFERENCES

[1] Abdelsamad S.F., Morsi W.G., and Sidhu T.S., 2015. Probabilistic impact of transportation electrification on the loss-of-life of distribution transformers in the presence of rooftop solar photovoltaic. *IEEE Transactions on Sustainable Energy* 6: 1565–73.

[2] Hopkins M. and A. Pahwa. 2008. Monte-Carlo simulation of energy production by a small wind generator. In *Proceedings of th 40th North American Power Symposium*, USA, pp.1–6.

[3] Sharma H., Bora B., Sharma S., Kumar R. and Jain R., 2017. Reliability study of solar pv power production in terms of weather parameters using Monte Carlo simulation. *International Journal of Engineering Research and Applications* 07: 37–45.

[4] Liu B.Y.H. and R.C. Jordan. 1960. The interrelationship and characteristic distribution of direct, diffuse and total solar radiation. *Solar Energy* 4: 1–19.

[5] Morf H., 2013. A stochastic solar irradiance model adjusted on the Ångström-Prescott regression. *Solar Energy* 87: 1–21.

[6] De Alfragide E., 1992. Tag : a time-dependent, autoregressive, Gaussian model for generating synthetic hourly radiation. *Solar Energy* 49: 167–74.

[7] Brecl K. and M. TopiI. 2014. Development of a stochastic hourly solar irradiation model. *International Journal of Photoenergy* 28(1): 33-40.

[8] Aguiar R.J., Collares-Pereira M. and Conde J.P., 1988. Simple procedure for generating sequences of daily radiation values using a library of Markov transition matrices. *Solar Energy* 40: 269–79.

[9] Graham V.A. and K.G.T. Hollands. 1990. A method to generate synthetic hourly solar radiation globally *Solar Energy* 44: 333–41.

[10] Suehrcke H. and P.G. McCormick. 1988. The frequency distribution of instantaneous insolation values. *Solar Energy* 40: 413–22.

[11] Besharat F., Dehghan A.A. and Faghih A.R., 2013. Empirical models for estimating global solar radiation: A review and case study. *Renewable and Sustainable Energy Reviews* 21: 798–821.

[12] Grantham A.P., Pudney P.J. and Boland J.W., 2018. Generating synthetic sequences of global horizontal irradiation. *Solar Energy* 162: 500–9.

[13] Ngoko B.O., Sugihara H., and Funaki T., 2014. Synthetic generation of high temporal resolution solar radiation data using Markov models. *Solar Energy* 103: 160–70.

[14] Bright J.M., Smith C.J., Taylor P.G., and Crook R., 2015. Stochastic generation of synthetic minutely irradiance time series derived from mean hourly weather observation data. *Solar Energy* 115: 229–42.

[15] Duffie J.A and W.A Beckman. 2013. *Solar engineering of thermal processes* (Wiley).

[16] Polo J., Zarzalejo L.F., Marchante R., and Navarro A.A., 2011. A simple approach to the synthetic generation of solar irradiance time series with high temporal resolution. *Solar Energy* 85: 1164–70.

[17] Grantham A.P., Pudney P.J., Ward L.A., Belusko M., and Boland J.W., 2017. Generating synthetic five-minute solar irradiance values from hourly observations *Solar Energy* 147: 209–21.

[18] Bright J.M., Babacan O., Kleissl J., Taylor P.G. and Crook R., 2017. A synthetic, spatially decorrelating solar irradiance generator and application to a LV grid model with high PV penetration. *Solar Energy* 147: 83–98.

[19] Azzouz M.A., Shaaban M.F., and El-Saadany E.F., 2015. Real-time optimal voltage regulation for distribution networks incorporating high penetration of PEVs. *IEEE Transactions on Power Systems* 30: 3234-3247.

[20] Gordon J.M. and T.A. Reddy. 1988. Time series analysis of daily horizontal solar radiation. *Solar Energy* 41: 215–26.

[21] Government of Canada 2015. Engineering Climate Datasets - *Canadian Weather Energy and Engineering Datasets (CWEEDS)* Retrieved on May 20, 2020 from the World Wide Web: https://climate.weather.gc.ca/prods_servs/engineering_e.html .

[22] Solar Anywhere 2019. *Solar Anywhere Datasets* Retrieved on May 20, 2020 from World Wide Web: https://www.solaranywhere.com/solutions/solaranywhere-data/.

[23] Perez R., Seals R., Ineichen P., Stewart R. and Menicucci D., 1987. A new simplified version of the perez diffuse irradiance model for tilted surfaces. *Solar Energy* 39: 221–31.

[24] Peel M.C., Finlayson B.L. and McMahon T.A., 2007. Updated world map of the Köppen-Geiger climate classification *Hydrology and Earth System . Sciences* 11: 1633–44.

[25] Peruchena C.F., Larrañeta M., Blanco M., and Bernardos A., 2018. High frequency generation of coupled GHI and DNI based on clustered dynamic paths. *Solar Energy* 159: 453–7.

[26] Larrañeta M., Moreno-Tejera S., Silva-Pérez M.A., and Lillo-Bravo I., 2015. An improved model for the synthetic generation of high temporal resolution direct normal irradiation time series. *Solar Energy* 122: 517–28.

[27] Munkhammar J., 2020. Polynomial Probability Distribution Estimation *MATLAB Central File Exchange*. Retrieved on May 20, 2020 from the World Wide Web: https://www.mathworks.com/matlabcentral/fileexchange/62520-polynomial-probability-distribution-estimation.

[28] R2017a - MATLAB & Simulink. *MathWorks* Retrieved on May 20, 2020 from World Wide Web: https://www.mathworks.com/products/matlab.html.

[29] Chamola V. and B. Sikdar. 2010. Synthetic generation of hourly solar irradiance using a multi-state Markov model. In Proceedings of the *IEEE Conference Publications* 1 1–3.