A Simulation Approach to Multi-station Solar Irradiance Data Considering Temporal Correlations

Xingbo Fu
Xi'an Jiaotong University

CONTENTS

Introduction

Solar irradiance

Uncertain

Intermittent

Meteorological Conditions

Cloud Amount

Clearness

Relative Humidity

Solar irradiance is one of the most significant factors that influence PV power generation.

Introduction

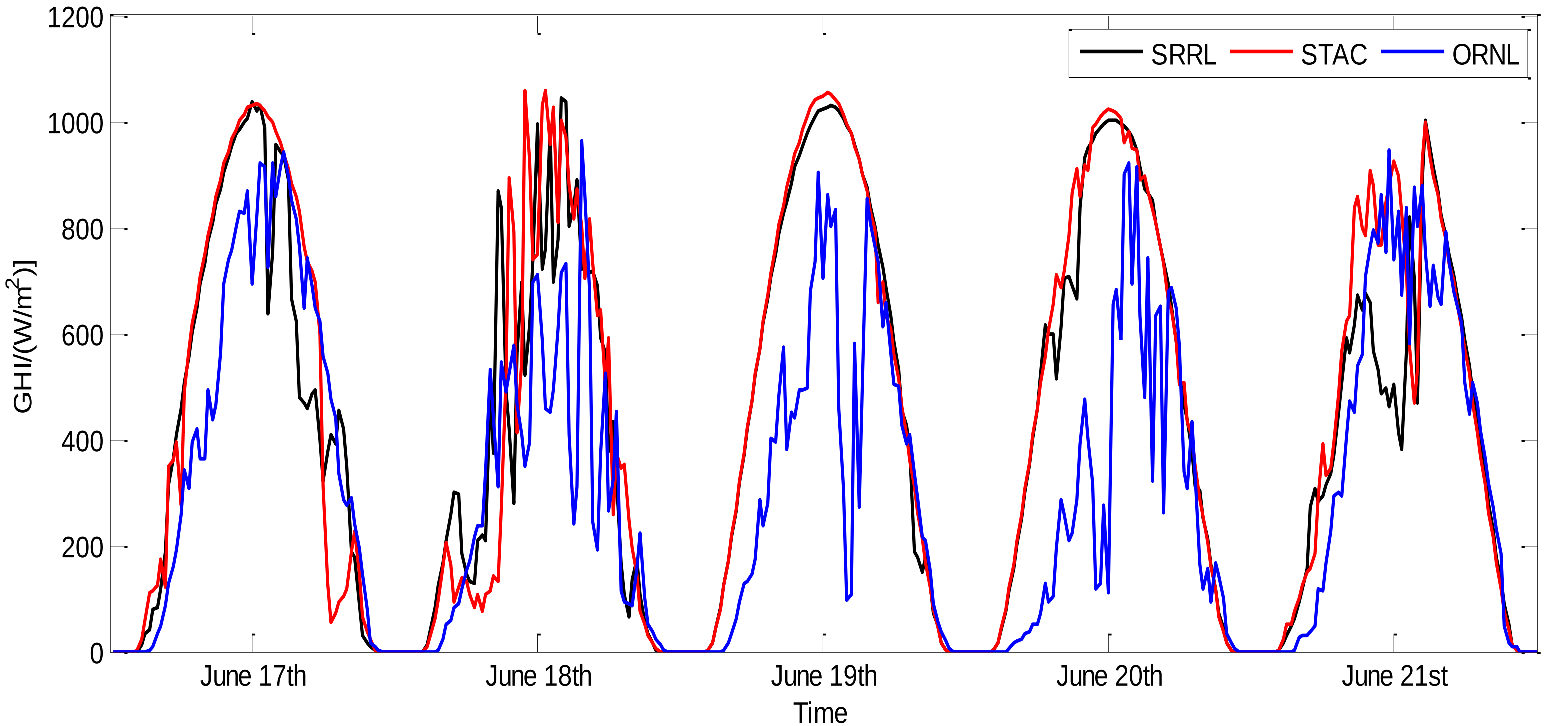
Solar irradiance data from National Renewable Energy Laboratory (NREL)[1] include one-minute-interval global horizontal irradiance (GHI)[2] data collected from many stations.

Details of Three Stations

Stations	Latitude	Longitude	Location
Solar Radiation Research Laboratory (SRRL)	39.74° North	105.18° West	Golden, Colorado
Solar Technology Acceleration Center (STAC)	39.76 ° North	104.62° West	Aurora, Colorado
Oak Ridge National Laboratory (ORNL)	35.93° North	84.31° West	Oak Ridge, Tennessee

1. https://midcdmz.nrel.gov
2. Total hemispheric shortwave irradiance as measured by an Kipp & Zonen Model with calibration factor traceable to the World Radiometric Reference (WRR)

Introduction



GHI Data of Three Stations (June 17th, 2015~ June 21st, 2015)

How to simulate GHI data of several stations?

Simulation Approach

A. Daily Features of GHI Data

As a kind of time series, one-minute-interval GHI data of one site consist of 1,440 daily samples. These samples are too large to be inputs for some algorithms like clustering so it is necessary to extract daily features of GHI data.

Standard Deviation → Amount of Dispersion

Skewness → Asymmetry

Simulation Approach

A. Daily Features of GHI Data

List of Daily Features

Feature Name	Description
Mean	$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x(i)$
Standard Deviation	$\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x(i)-\bar{x})^2}$
Skewness	$\gamma = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{x(i)-\bar{x}}{\sigma}\right)^3$
Kurtosis	$\beta = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{x(i)-\bar{x}}{\sigma}\right)^4 - 3$
Moving Fluctuation Intensity	$\mathrm{MFI} = \mathrm{RFC} \times \mathrm{AFM}$

IEEE PES Power & Energy Society®

2019 ISGT Innovative Smart Grid Technologies Asia

Simulation Approach

A. Daily Features of GHI Data

Moving Fluctuation Intensity (MFI) is defined as

$$MFI = RFC \times AFM$$

Reverse Fluctuation Count (RFC) is defined as count value that when fluctuation trend is reversed, then count value add 1. Fluctuation trend is reversed when

$$[x(i+1) - x(i)][x(i) - x(i-1)] < 0$$

Average Fluctuation Magnitude (AFM) is defined as

$$AFM = \frac{1}{N}\sum_{i=1}^{N} abs[x(i+1) - x(i)]$$

Simulation Approach

B. K-means Clustering

For station i, feature vectors f_i of n-day historical GHI data

$$f_i = (f_i^1, f_i^2, \dots, f_i^m \dots, f_i^{n-1}, f_i^n)$$

K-means Clustering groups all the days to k clusters $C = \{c_1, c_2, \dots, c_k\}$ based on their normalized feature vectors of GHI data. Each cluster is defined as a state of PV generation suitability.

For station i, states s^i of n-day historical GHI data

$$s^i = (s_1^i, s_2^i, \dots, s_t^i, \dots, s_n^i)$$

Simulation Approach

B. K-means Clustering

State matrix s of all these j stations

$$s = \begin{pmatrix} s^1 \\ s^2 \\ \vdots \\ s_i \\ \vdots \\ s^j \end{pmatrix} = \begin{pmatrix} s_1^1, s_2^1, \ldots, s_t^1, \ldots, s_n^1 \\ s_1^2, s_2^2, \ldots, s_t^2, \ldots, s_n^2 \\ \vdots \\ s_1^i, s_2^i, \ldots, s_t^i, \ldots, s_n^i \\ \vdots \\ s_1^j, s_2^j, \ldots, s_t^j, \ldots, s_n^j \end{pmatrix}$$

s can be written as

$$s = (s_1, s_2, \cdots, s_t, \cdots, s_n)$$

s_t is states of the whole j stations on day t

$$s_t = \left(s_t^1, s_t^2, \cdots, s_t^i, \cdots, s_t^j\right)^T$$

Simulation Approach

B. K-means Clustering

When we group days to k clusters, there are $r = k^j$ possible state permutations. We convert this state space into digital code.

Digital Code	Reduced State Space
1	$(c_1, c_1, \cdots, c_1)$
2	$(c_1, c_1, \cdots, c_2)$
$\vdots$	$\vdots$
$r-1$	$(c_k, c_k, \cdots, c_{k-1})$
r	$(c_k, c_k, \cdots, c_k)$

$$s = (s_1, s_2, \cdots, s_t, \cdots, s_n) \rightarrow s = (d_1, d_2, \cdots, d_t, \cdots, d_n)$$
$$d_t = 1,2, \cdots, r$$

Simulation Approach

C. MTPM

A Markov chain is a type of Markov process that has a discrete state space. This stochastic process that satisfies the Markov property - the future state d_{t+1} only depends on the present state d_t and it does not depend on the previous state $d_1, d_2, \cdots, d_{t-1}$.

$$P\{s(t+1) = d_{t+1} | s(1) = d_1, s(2) = d_2, \cdots, s(t) = d_t\}$$
$$= P\{s(t+1) = d_{t+1} | s(t) = d_\mathrm{t}\}$$

Simulation Approach

C. MTPM

In order to describe the changes of states, Markov chain $\{d_1, d_2, \cdots, d_t, \cdots, d_n\}$ can be characterized by the first-order Markov Transition Probability Matrix (MTPM) P consisting of the transition probabilities

$$P = \begin{pmatrix} p_{11} & \cdots & p_{1r} \\ \vdots & \ddots & \vdots \\ p_{r1} & \cdots & p_{rr} \end{pmatrix}$$

$$p_{ij} = P\{s(t+1) = j | s(t) = i\}$$

$$for\ i, j = 1,2, \cdots, r$$

Simulation Approach

D. State Space Reduction

Not all these state permutations happen in the real world. Hence, state space reduction can be applied before calculating MTPM so that the dimensions of MTPM can be reduced. It means that if the state permutation s_t does not happen in n-day historical GHI data, we will cut off this state permutation from state space. When r_0 different state permutations do not happen, the size of reduced state permutations are

$$r = k^j - r_0$$

Simulation Approach

E. Simulation Approach

Step 1: Set the initial random daily state s_{init} and size of days to simulate n. $s' = \{ s_{init} \}$.

Step 2: Given a random number $r \in (0,1)$, find out the next daily state s_{next}, which satisfies

$$p_{s_{init}\ s_{next}} < r < \sum_{i=1}^{s_{next}} p_{s_{init}\ i}$$

Step 3: Append s_{next} to s' and let $s_{init} = s_{next}$.

Step 4: When the size of elements in s' reaches n, go to Step 5. Otherwise, return to Step 2.

Step 5: Convert s' into state matrix s according to Table III.

Step 6: As for s_t^i in s, select one day whose state is s_t^i from the observed days.

Step 7: Connect the raw data of days from Step 6 to get the simulated data.

Case Study

A. GHI Data

We collect the observed one-minute-interval GHI data from January 1st, 2012 to December 31st, 2018. Considering different daily sunshine durations among different seasons, we divide days of the observed GHI data into four seasons.

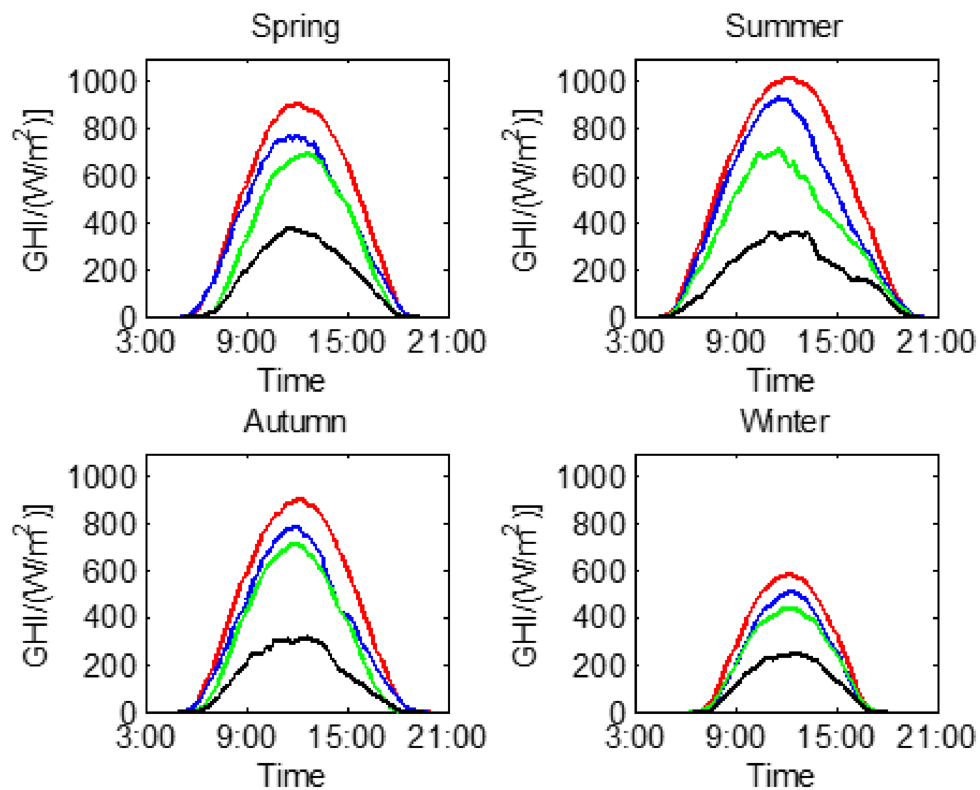
Seasons	Size of Days
Spring	625
Summer	644
Autumn	644
Winter	644

Case Study

B. Results of Clustering

We group GHI data into four clusters for different seasons at SRRL, STAC and ORNL stations.

The red curves are the most suitable for generation while the black the worst. Therefore, the red curves represent c_1, the blue curves represent c_2, the green curves represent c_3 and the black curves represent c_4.



Average STAC's GHI data from the Same Cluster

Case Study

C. Performance of Simulation

Statistical Analysis - Compare some statistical properties (mean and standard deviation) of the simulated one-year GHI data with those of the observed GHI data.

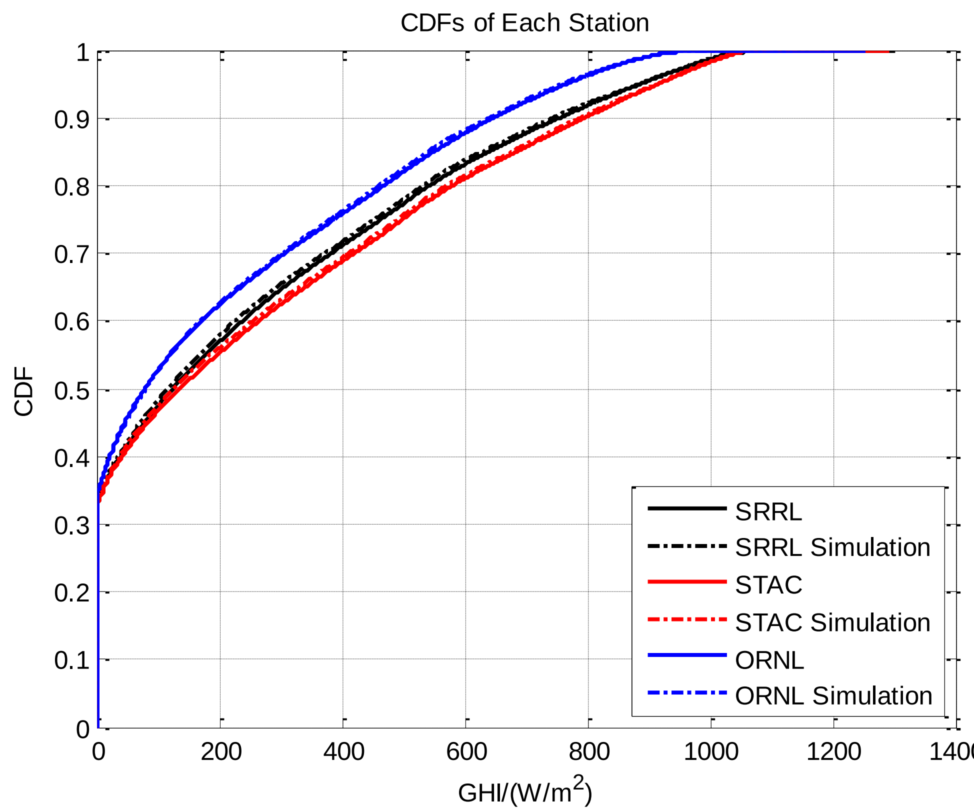
Statistical Properties of Simulated Data and Observed Data

Year	SRRL		STAC		ORNL	
	Mean	Std	Mean	Std	Mean	Std
2012	265.63	270.56	284.01	285.43	226.39	233.43
2013	261.12	268.44	278.28	282.39	208.68	219.92
2014	253.42	262.10	270.03	276.04	220.99	229.38
2015	249.86	259.60	270.13	276.96	212.61	219.72
2016	265.30	273.03	282.77	285.26	227.32	233.60
2017	254.76	263.61	268.42	274.58	206.31	213.24
2018	260.36	267.30	272.40	276.65	193.79	201.04
All	258.64	266.38	275.16	279.62	213.74	221.49
Simulation	260.76	266.15	281.26	282.34	211.27	221.61

Case Study

C. Performance of Simulation

Cumulative distribution function (CDF) – CDF describes the cumulative probability associated with a random variable and the approximate CDFs between the observed GHI data and simulated data is required.

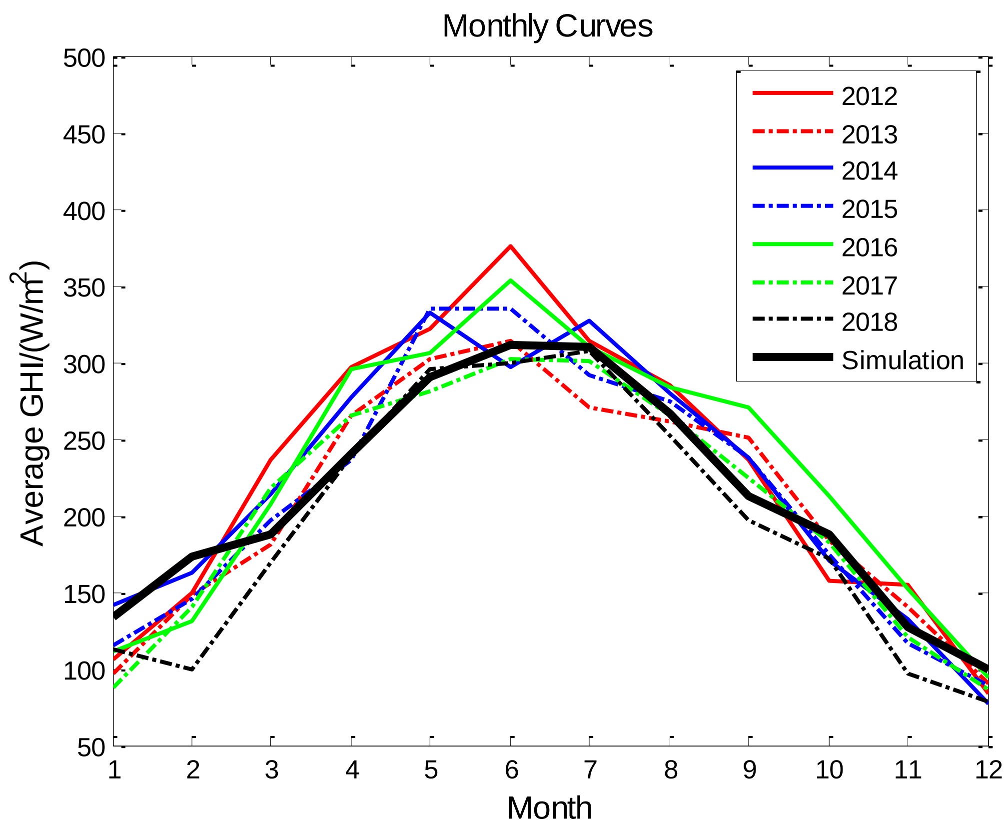


CDFs of Simulated Data and the Observed Data

Case Study

C. Performance of Simulation

Monthly curves of the observed GHI data every year and the simulated data.



ORNL Monthly Curves

Thank you!

Xingbo Fu
Xi'an Jiaotong University