# Synthetic Generation of Hourly Solar Irradiance Using a Multi-State Markov Model

Vinay Chamola and Biplab Sikdar Department of Electrical and Computer Engineering National University of Singapore, Singapore Email: {vinay.chamola,bsikdar}@nus.edu.sg

*Abstract*—Solar irradiance data is a key factor in dimensioning photovoltaic (PV) panel size and energy storage for solar powered systems. Unfortunately, long term solar irradiace data is not available for many locations, thereby affecting the accuracy and effectiveness of dimensioning efforts. To address this issue, this paper presents a model for generating accurate, synthetic traces of arbitrary length for hourly solar irradiance. The proposed model is based on an approach that combines daily correlations in weather conditions affecting the solar irradiance levels. The proposed model has been evaluated by comparing various statistical parameters of hourly solar irradiance against real data.

## I. INTRODUCTION

With the increase in global carbon emissions and the increasing awareness in government agencies and industries towards environmental issues, green energy solutions have emerged as a solution and active area of research in many domains such as transportation, communications, electrification etc. This has led to popularity in the use of solar energy for powering various systems. One of the key steps in dimensioning the PV panel and battery size for such applications is to accurately model the solar irradiation at the location [1].

Popular existing approaches to model the daily/hourly solar irradiance include Markov models [2] and autoregressive moving average (ARMA) models [3]. The Markovian approach assumes that the solar irradiance at a given time (either day or hour) is dependent only on the solar irradiance at the last time unit. Further, daily or hourly transition statistics are derived to characterise the transition from one solar irradiance level to other. The ARMA approach uses autocorrelation between consecutive days to capture the nature of the solar radiation. It has been shown that Markov models perform better than ARMA models [4]. While most Markov models focus on daily irradiance levels, some models exist for modeling the hourly solar irradiation [4]. However, these hourly models lack accuracy since they do not consider the day to day correlations in the solar irradiance. To address this issue, this paper proposes a Markov model which captures the solar irradiance characteristics both on an hourly as well as on a daily scale.

# II. PROPOSED MODEL

In the proposed model, we classify any given day as either a "bad" or "good" weather day based on the daily solar irradiance level of that particular day. From the complete set



Fig. 1. a. An example of the demarcation of states and their expected solar irradiace for bad days. b. Markov state diagram showing transition from a state in an hour to state in next hour.

of days we sort the days based on their daily solar irradiation and we select  $\alpha\%$  of the days with lowest irradiation as bad days. All other days are classified as good days. Given a bad or good day, the next day may be either a bad or a good day. In order to capture the occurrence of consecutive bad or good weather days, we model the transition between days as a twostate Markov process. The transition probability matrix of this Markov process can be expressed as

$$\mathbf{T} = \begin{bmatrix} p_{bb} & p_{bg} \\ p_{gb} & p_{gg} \end{bmatrix}$$
(1)

where  $p_{bb}$  (respectively,  $p_{gg}$ ) is the probability of transition from a bad day (good day) to a bad day (good day), and  $p_{bg} = 1 - p_{bb}$  (respectively,  $p_{gb} = 1 - p_{gg}$ ) is the probability of transition from a bad day (good day) to a good day (bad day).

Now, in order to capture the hourly variations of the solar irradiance, each hour is categorized into one of the four possible states. The four states for bad and good days are different and each category is characterized by its solar irradiance level. For demarking the boundary between the categories, we find the hourly minimum and maximum for a day type and divide the region between them uniformly into four regions (as shown in Figure 1(a)). Also, the average value in each of these regions is computed and is used in our model as the representative solar irradiance value in that state. The overall state space of the process representing the solar irradiance at any hour is denoted by

$$S_t: S_t \in \{B_{x,y}, G_{x,y}\}, \quad x \in \{1, 2, \cdots, 24\}, y \in \{1, 2, 3, 4\}$$
(2)

where B and G correspond to bad and good weather days respectively, x denotes the hour of the day and y denotes the state of the solar irradiance in that hour. Based on the Markovian assumption for the hourly solar irradiance, the state of solar irradiance at a given hour depends only on the state of solar irradiance in the previous hour and the transition probability of going from the previous state to the current state. This can be expressed as:

$$P[S_t|S_{t-1}, S_{t-2}, \dots S_0] = P[S_t|S_{t-1}].$$
(3)

Given that the solar irradiance is currently in a given state, the state may transition to any of the four states in the next hour (for both good and bad weather days). For a bad weather day, the transition probability matrix can be given as

$$\mathbf{B} = \begin{bmatrix} b_{(1,1)(1,1)} & \cdots & b_{(1,1)(24,4)} \\ \vdots & \ddots & \vdots \\ b_{(24,4)(1,1)} & \cdots & b_{(24,4)(24,4)} \end{bmatrix}$$
(4)

with

$$b_{(i,j)(k,l)} = \begin{cases} r_{(i,j)(k,l)} & k = (i+1) \mod 24; \ j,l \in \{1,2,3,4\} \\ 0 & \text{otherwise} \end{cases}$$
(5)

where  $b_{(i,j)(k,l)}$  is the probability of transition from the *j*-th state in the *i*-th hour to the *l*-th state in *k*-th hour on a bad weather day. Note that in a given hour, from a particular state, the solar irradiance can only go to one of the states in the next hour. The numerical value of the transition probability is denoted by  $r_{(i,j)(k,l)}$  and it satisfies

$$\sum_{l=1}^{4} r_{(i,j)(k,l)} = 1, \qquad k = (i+1) \bmod 24, j \in \{1, 2, 3, 4\}.$$
(6)

We can similarly define the the transition probability matrix for a good weather day. The hourly state transitions for a bad day are shown in Figure 1(b).

#### **III. RESULTS**

To validate the methodology presented in this paper, we consider year 2009 for the location Kolkata (India) and generate a synthetic trace for one year using our methodology. In our model we use  $\alpha = 0.2$ . The solar irradiance data for the location was obtained from the NREL database [5].

# A. Daily Mean and Variance

To ascertain the accuracy of the proposed model, we first consider the mean and variance in the hourly solar irradiance values, calculated on a monthly basis for the year 2009 for Kolkata (Figure 2).We note the close match between the overall trends and values of the data for the synthetic and empirical traces.



Fig. 2. Comparison of the mean and variance of the daily solar irradiance values, empirical vs model generated data: Kolkata (2009).



Fig. 3. Comparison of number of hours in a particular irradiance slot, empirical vs model generated data: Kolkata (June 2009).

## **B.** Irradiation State Statistics

Next we consider the statistics related to the hourly irradiance levels. From Figure 2, we observe that the month of June for Kolkata has the highest values of variance in daily solar irradiation. This indicates that this is the month which exhibits maximum variability in terms of solar irradiance. Thus we pick this month to further test the accuracy of our model. For this month, using our model, we generate a month's worth of synthetic data. For both the empirical and the synthetic data this month, we count the occurrences of the hourly solar irradiance being in a particular irradiance slot. For this comparison we have divided the solar irradiance from 0 to  $1000W/m^2$  into 10 slots. The comparison is shown in Figure 3. It can be observed that there is a close match between the number of hours in a given irradiance slot for the model generated synthetic trace as compared to that obtained from empirical data.



Fig. 4. Comparison of hourly mean and variance, empirical vs model generated data: Kolkata (June 2009).

# C. Hourly mean and variance

To further evaluate the accuracy of the model, for the month of June 2009 (which exhibits maximum variability in the year 2009), we next compare the mean and variance of the solar irradiance in each of the hours of the day. Figure 4 shows the hourly mean and variance values for the empirical data and our proposed model. It can be seen that the hourly mean and variance of the synthetic trace generated by our model closely match with the hourly mean and variance of the empirical data.

## **IV. CONCLUSION**

This paper presented a framework for generating synthetic traces of hourly solar irradiance for the purposes of dimensioning solar-powered systems. The proposed framework is based on combining Markov models for the daily changes in weather patterns with one for an hourly variation in the solar irradiance levels. The proposed model has been validated by comparing it against empirical data.

# REFERENCES

- Zhou, Wei, et al. "Current status of research on optimum sizing of standalone hybrid solarwind power generation systems." Applied Energy 87.2 (2010): 380-389.
- [2] Poggi, Philippe, et al. "Stochastic study of hourly total solar radiation in Corsica using a Markov model." International journal of climatology 20.14 (2000): 1843-1860.
- [3] Festa, R., S. Jain, and C. F. Ratto. "Stochastic modelling of daily global irradiation." Renewable energy 2.1 (1992): 23-34.
- [4] Kamal, Lalarukh, and Yasmin Zahra Jafri. "Stochastic modeling and generation of synthetic sequences of hourly global solar irradiation at Quetta, Pakistan." Renewable energy 18.4 (1999): 565-572.
- [5] http://www.nrel.gov/rredc/solar\_data.html, Last accessed: August 20, 2014.