Available online at www.sciencedirect.com
ScienceDirect
Solar Energy 103 (2014) 160–170
SOLAR ENERGY
www.elsevier.com/locate/solener

Synthetic generation of high temporal resolution solar radiation data using Markov models

B.O. Ngoko *, H. Sugihara, T. Funaki

Division of Electrical, Electronics and Information Engineering, Osaka University, 2-1 Yamadaoka, Suita, Osaka 565-0871, Japan

Received 19 February 2013; received in revised form 2 February 2014; accepted 11 February 2014
Available online 3 March 2014

Communicated by: Associate Editor David Renne

Abstract

In this paper, a model for the synthetic generation of 1-min global solar radiation data starting from the daily clearness index is presented. The model is constructed by treating the process generating a normalized form of the 1-min clearness index sequence as a Markov process. Two sets of three-year global solar radiation data taken at 1-min intervals from two locations in Japan are used to construct the Markov transition matrices. Because different days have different statistical characteristics due to different weather conditions, the days in the data set are first clustered into groups based on the daily clearness index values. Transition matrices are then formed for each group and consequently used to synthetically generate 1-min global solar radiation data. Second-order Markov models are selected based on the partial autocorrelation functions of the measured data. The statistical characteristics of the measured and synthetic data sets are found to be in close agreement thus confirming the validity of the model.

Keywords: Solar power; Markov model; Synthetic generation; Probabilistic studies

1. Introduction

The REN21 (2011) global status report on renewable energy cites solar power as one of the fastest growing sources of electricity generation. In the 2009–2010 period alone, the total installed solar photovoltaic capacities increased by more than 70% from 23 GW to 40 GW globally. A large portion of this capacity is installed in the form of micro-generators such as rooftop PV generators installed in power distribution systems (Liu et al., 2010). As the penetration levels of such generators in power systems increases, it is important to study what effects they would have on the system. Such studies are important both for planning and system operations purposes and more so because of the highly variable nature of solar power (Tan and Kirschen, 2007; Dahal et al., 2010).

While deterministic studies may be used to draw general conclusions about the effects of increased integration of solar generators in power systems, especially in distribution systems, such studies do not take into account the variable nature of solar power. On the other hand, probabilistic studies may be used to capture this characteristic, hence can provide much more detailed analysis. Such probabilistic studies can be carried out using solar irradiation as an input to the simulation program. One approach would be to use recorded solar radiation data. However, historical solar radiation data taken at high sampling frequencies are unavailable for many locations thus limiting the scope of researchers to specific sites where such data exists. As a different approach, it may be preferable to use computer simulated data as long as the model used to synthetically

* Corresponding author. Tel.: +81 06 68797709; fax: +81 06 68797713.
E-mail address: ngokob@ps.eei.eng.osaka-u.ac.jp (B.O. Ngoko).

http://dx.doi.org/10.1016/j.solener.2014.02.026
0038-092X/

generate the data produces data of similar statistical characteristics to typical measured data.

A number of models for global solar radiation have been proposed in literature. Most of these models have been constructed for daily or hourly radiation (Graham et al., 1988; Aguiar et al., 1988; Graham and Hollands, 1990; Aguiar and Collares-Pereira, 1992; Palomo, 1989) mainly because of availability of such data. The proposed models use either Autoregressive Moving Average (ARMA) or Markov Transition Matrix (MTM) methods to model the clearness index – a ratio of the solar radiation recorded on the earth's surface to the extraterrestrial solar radiation. However, the statistical properties of the daily or hourly clearness index data are significantly different from the statistical properties of 1-min or even 5-min clearness index data (Jurado et al., 1995; Suehrcke and McCormick, 1988). Both Jurado et al. (1995) and Suehrcke and McCormick (1988) report bimodal probability density functions of 1-min and 5-min clearness index data; a feature significantly different from the unimodal nature of the probability density functions of daily or even hourly clearness index data. These differences imply that models for various temporal resolutions would be different.

While most of the studies on probabilistic modeling and synthetic generation of solar radiation data have been focused on daily and hourly data, there have been attempts at modeling higher temporal resolution solar radiation data (Skartveit and Olseth, 1992; Richardson and Thomson, 2011). Skartveit and Olseth (1992) model the probability distribution and lag-1 autocorrelation of short term irradiance data (1–10 min) and use these in a first order autoregressive model for the synthetic generation of short term data. While the models achieve satisfactory results in reproducing the modeled data, they utilize complex approximate equations of the probability density functions and the errors introduced by these approximations are not quantified. The model also ignores the dependence of clearness index on solar elevation angle as reported by Perez et al. (1990). Also, by only considering the lag-1 autocorrelation, the model assumes little or no correlation in the time series data to data at lags greater than 1.

Richardson and Thomson (2011) use a single first order Markov Transition Matrix to synthetically generate 1-min global solar radiation data. The model is a single part of a large integrated model which also includes a household occupancy and power demand model. Though it is quite simple, the use of a single transition matrix ignores the fact that different days have different statistical properties. For example, it is expected that a transition matrix modeling a clear day will be significantly different from one modeling a cloudy day.

A Markov model for the synthetic generation of 1-min global solar radiation data is presented in this paper. Because of the dependence of clearness index on solar elevation, a normalized form of the clearness index is modeled. Also, based on a realization that different days have different statistical characteristics, the days in the data set are first grouped based on the daily clearness index value and parameters for the Markov models are then extracted for each group. A look at the nature of the autocorrelation characteristics of measured data leads to the choice of a second order Markov model.

The proposed model is validated by comparing the statistical characteristics of the synthetically generated data to those of the observed data. The statistical characteristics compared in the model validation include the ordinary moments, probability distributions, autocorrelation functions, and minute-to-minute irradiation fluctuations. A close agreement is found between the statistical characteristics of the synthetic and observed data sets.

2. Markov models

Markov models provide a simple yet powerful way of modeling the dependence between adjacent observations in a given time series. They have consequently been used extensively in the modeling of various stochastic variables including the modeling of wind time series (Shamshad et al., 2005) and rainfall patterns (Srikanthan and McMahon, 1985). They have also been used in the modeling of global solar radiation data (Aguiar et al., 1988; Palomo, 1989).

A process is said to exhibit the Markov property if, given its present state, the future is conditionally independent of the past and Markov models are mathematical representations of such stochastic processes (Häggström, 2002). A Markov process (X_t, $t = 0, 1, 2, \ldots$) with a set of m allowed states $(1, 2, \ldots, m)$ is said to be in state j at time t if $X_t = j$. In a first order Markov process, given that the process is in state i at time $t-1$, the probability that it will be in state j at time t is given by a fixed probability P_{ij} written mathematically as:

$$\begin{aligned} P_{ij} &= P(X_t = j | X_{t-1} = i, X_{t-2} = i_{t-2}, \ldots, X_0 = i_0) \\ &= P(X_t = j | X_{t-1} = i). \end{aligned} \tag{1}$$

P_{ij}, known as the transition probability from state i to j, is independent of the states of the process at times $t-2, t-3, \ldots$. This is known as the memoryless property of Markov models – the conditional distribution of X_t given $X_0, X_1, \ldots X_{t-1}$ depends only on X_{t-1}.

In an nth order Markov model, the probability that the process will be in a particular state at time t depends not only on its state at time $t-1$ but also on the states at times $t-2, t-3, \ldots, t-n$. For example, in a second order Markov process, the probability that the process will be in state k at time t given that it was in state j at time $t-1$ and in state i at time $t-2$ is given by:

$$P_{ijk} = P(X_t = k | X_{t-1} = j, X_{t-2} = i). \tag{2}$$

The transition matrix $\mathbf{P^n}$ holds the transition probabilities for the nth order Markov process. For a process with m allowed states, the first order transition matrix would typically be represented as: