Solar Energy 170 (2018) 174–183

Contents lists available at ScienceDirect

Solar Energy

journal homepage: www.elsevier.com/locate/solener

A Markov-chain probability distribution mixture approach to the clear-sky index

Joakim Munkhammar*, Joakim Widén

Built Environment Energy Systems Group, Department of Engineering Sciences, Uppsala University, SE-751 21 Uppsala, Sweden

ARTICLE INFO

Keywords:
Clear-sky index
Markov-chain modeling
Probability distribution modeling
Autocorrelation function

ABSTRACT

This paper presents a Markov-chain probability distribution mixture approach to the clear-sky index (CSI). The main assumption is that the temporal variability of the state of clear and the state of cloudy can be described by a two-state Markov-chain, and the variability within each state can be approximated by a probability distribution, unique for each state. Measurables such as the mean clear-sky index, fraction of bright sunshine, expected duration of clearness and expected duration of cloudiness events are shown to be related to the parameters of the method. Additionally, the Ångström equation, which relates mean normalized solar irradiance to the fraction of bright sunshine, is shown to arise as the expectation of the method. In order to numerically verify the method, a simulation model is constructed based on data sets for two different climatic regions: Norrköping, Sweden and Oahu, Hawaii, USA. Results from the simulation model based on training data shows good agreement with testing data, and when comparing the results to existing models in the literature it is comparable to the state of the art. It is shown that the simulation model generates a non-trivial, generally non-zero, autocorrelation function. Finally, challenges with the method and open problems are discussed.

1. Introduction

Solar irradiance has interesting statistical temporal features, particularly on minute to instantaneous scale (Bright et al., 2015), and for such resolution the normalized clear-sky index distribution can be described as a probability distribution with typically two or three peaks (Hollands and Huget, 1983; Suehrcke and McCormick, 1988; Hollands and Suehrcke, 2013; Munkhammar et al., 2015a; Widén et al., 2017). Studies on clear-sky index temporal variability include using Markov-chains (Bright et al., 2015), neural networks (Voyant et al., 2011), copulas (Munkhammar and Widén, 2016), bootstrapping (Grantham et al., 2018) and pure sampling from probability distributions (Munkhammar et al., 2015a,b). Input data varies in complexity from utilizing cloud size, coverage and morphology (Bright et al., 2015; Smith et al., 2017) to input data of the clear-sky index time-series (Munkhammar et al., 2015a,b; Munkhammar et al., 2017a; Munkhammar and Widén, 2017a).

A challenge is to not only fit an accurate probability distribution to the clear-sky index, but to configure a model which generates realistic synthetic temporal variability as well (Bright et al., 2017). Time-series realism is typically measured via the autocorrelation function (Munkhammar and Widén, 2017a; Brinkworth, 1977; Skartveit and Olseth, 1992). Autocorrelation of the clear-sky index time-series has been studied previously for instantaneous irradiance (Munkhammar and Widén, 2017a; Brinkworth, 1977; Skartveit and Olseth, 1992; Aguiar and Collares-Pereira, 1992; Hansen et al., 2010; Munkhammar and Widén, 2017b), where the autocorrelation function is positive and follows an exponential slope for hour resolution (Skartveit and Olseth, 1992; Aguiar and Collares-Pereira, 1992; Hammer and Beyer, 2013), while it has also shown negative values for minute resolution (Munkhammar and Widén, 2017a,b; Hansen et al., 2010).

Models for generating synthetic clear-sky index data utilizing the autocorrelation of the clear-sky index include Gaussian-Markov (Brinkworth, 1977), auto-regressive Gaussian (Aguiar and Collares-Pereira, 1992), neural networks (Voyant et al., 2011), copula modeling (Munkhammar and Widén, 2017a,b) and fractal cloud modeling (Lohmann et al., 2017). There are also Markov-chain models for modeling solar irradiance, where (Bright et al., 2015; Morph, 1998; Aguiar et al., 1988; Palomo, 1989; Ngoko et al., 2014) are canonical examples. These models are examples of so-called clear-sky or clearness index generators.

Clear-sky or clearness index generators are practical, since they use some existing data set of, e.g., lower resolution or averaged clear-sky index or clearness data to estimate higher resolution data (temporal or spatial), see e.g. (Bright et al., 2015, 2017; Morph, 1998; Ngoko et al., 2014; Wegener et al., 2012; Grantham et al., 2017). Markov-chains for

* Corresponding author.
E-mail addresses: joakim.munkhammar@angstrom.uu.se (J. Munkhammar), joakim.widen@angstrom.uu.se (J. Widén).

https://doi.org/10.1016/j.solener.2018.05.055
Received 27 February 2018; Received in revised form 7 May 2018; Accepted 14 May 2018
0038-092X/

solar irradiance has in general been used to create clear-sky index generators, see (Morph, 1998; Aguiar et al., 1988; Palomo, 1989; Ngoko et al., 2014), where the Markov-chain induces non-zero autocorrelation. Here (Ngoko et al., 2014) generated minute resolution, while (Aguiar et al., 1988) generated daily resolution. In Morph (1998) a general model for generating clear and cloudy periods was devised. Generally, it should be emphasized that models for generating synthetic clear-sky index data are complements to irradiation estimates from measurements, e.g. from satellite data, see e.g. (Engerer et al., 2017), and software, such as Meteonorm (irradiance software, 2017).

A conclusion from the literature is that there is a challenge in designing as parsimonious mathematical structures as possible for describing the clear-sky index, while maintaining as high goodness-of-fit as possible when compared with measured data and other studies. The temporal variability in the existing modeling of the clear-sky index is generally overwhelmingly complex, where typical examples are Ngoko et al.'s and Bright et al.'s Markov-chain approaches (Bright et al., 2015; Ngoko et al., 2014). The example presented in (Morph, 1998) is a good example of a simple model with complex dynamics.

This study aims to develop a method and a simulation model based on the method, for investigating the temporal variability of the clear-sky index. The method and the model are intended to be as close to state of the art as possible in terms of providing ample statistical information, while also being parsimonous in its mathematical formulation.

The method is a continuation of the original statistical method development in Ångström (1922) and Hollands and Suehrcke (2013), the modeling in Ngoko et al. (2014) and Widén et al. (2017). It is a continuation of the concept of a two-state Markov-chain model in (Morph, 1998) by separating sunny and cloudy periods with a Markov-chain, but presents different dynamical structure for the states. Both the method and the model that are presented in this study are new to the solar research community. The method and the model represent the first step of using Markov-chains in mixture with probability distribution for analyzing the clear-sky index.

The model is a clear-sky index generator, similar to the models of Grantham et al. (2018, 2017), Ngoko et al. (2014) and Munkhammar and Widén (2017a,b). Although also Markov-chain based, the method and the model in this study require less input than the Ngoko et al. model, which use hourly weather observations including sea level pressure, wind speed, cloud base height and cloud cover (Ngoko et al., 2014). Also, only global horizontal irradiance (GHI) is used, and not direct normal irradiance (DNI), like in (Grantham et al., 2017). The method and the model are less mathematically and computationally complex than for example the copula-based model in (Grantham et al., 2018; Munkhammar and Widén, 2017a,b).

Solar irradiance data sets for Norrköping and Hawaii are used in the simulations, which were also used in (Munkhammar and Widén, 2017a).

This paper is organized as follows. In Section 2 the methodology for the method, the simulation model and data are presented. In Section 3 the results from the simulation model are presented, in Section 4 the conclusions are presented and in Section 5 the results are discussed in a wider context with particular focus on open problems.

2. Methodology

The methodology is organized so that in Section 2.1 the main method is presented. In Section 2.3 the data set is presented and in Section 2.2 the goodness-of-fit statistic used in the simulation model is presented. In Section 2.4 the simulation model is presented.

2.1. Markov-chain probability distribution mixture statistics of the clear-sky index

The following observations can be made regarding the state of solar irradiance at any time, which is similar to an assumption made in (Widén et al., 2017):

- Instantaneous solar irradiance is characterized by two mutually exclusive states where either is occupied at any time of day: clear and cloudy.
- There is a stochastic variability in solar irradiance within each state.

These observations will be defined as *first and second principles of solar radiation variability* in this study. The first principle is based on separating the mutually exclusive states of clear and cloudy, as was similarly modeled in (Widén et al., 2017), although for spatial solar irradiance variability. The second principle is a crude approximation of the variability in the states as a guiding principle for the parsimonious approach to the clear-sky index in this study. In order to define and delimit the method developed, two approximations of the dynamics of solar radiation will be assumed:

A. The persistence over time in each state and the probability of transition between states of clear and cloudy is expressed as a two-state Markov-chain.
B. The variability of the clear-sky index in each of the states is approximated by a probability distribution model for each state.

Approximation A is motivated by the semi-persistent variability of clear and cloudy states; solar irradiance data indicates that there is a higher resolution variability in the solar irradiance within in each of clear and cloudy states respectively, at least on minute resolution or higher. Approximation B is the simplest possible model for the variability within each of the two states, similar to probability distribution fitting of the clear-sky index like in e.g. (Hollands and Huget, 1983; Hollands and Suehrcke, 2013; Munkhammar et al., 2015a; Widén et al., 2017). The idea with this study is to find a method which has reasonable accuracy in terms of probability distribution and temporal variability with a minimum of additional statistical structure.

Schematics illustrations of the Markov-chain probability distribution mixture method (including transition graph) and the simulation model is shown in Fig. 1. The formal statistical setup of the method follows.

This study focuses on the indeterministic temporal variability of instantaneous solar irradiance, the clear-sky index, defined as the variability of the solar irradiance where the deterministic variability of the sun's position on the sky dome has been removed.

Formally, the clear-sky index κ is defined as the ratio between the measured global horizontal irradiance (GHI) $G(t)$ and the estimated global horizontal clear-sky irradiance $G_c(t)$ over time t:

$$\kappa(t) \equiv \frac{G(t)}{G_c(t)}. \tag{1}$$

For the method developed in this paper, let the clear-sky index κ be approximated by the following construction, similar to the spatial model in Widén et al. (2017):

$$\kappa(t) = X_t\kappa_c + (1-X_t)\kappa_s \tag{2}$$

where κ_c and κ_s are yet undefined independent stochastic variables. X_t is a Markov-chain of two states with outcomes $X_t \in [0, 1]$ for time t (Breuer and Baum, 2005, p.12,Cinlar, 1975, p.44). This implies, for the Markov-chain, that $X_t = 0$ holds for being in the clear state, and $X_t = 1$ represents the cloudy state. The clear state renders Eq. (9) $\kappa(t) = \kappa_s$, and for the cloudy state, Eq. (9) becomes $\kappa(t) = \kappa_c$.

This construction fulfills the first and second principles of solar irradiance variability and the approximations A and B, respectively, as well. Mathematically the Markov-chain describes the probability of transition P_{ij} from state $i \in [0, 1]$ to state $j \in [0, 1]$ over time t:

$$P_{ij} = P(X_t = j|X_{t-1} = i). \tag{3}$$