Solar Energy 158 (2017) 9–19

Contents lists available at ScienceDirect

Solar Energy

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

An autocorrelation-based copula model for generating realistic clear-sky index time-series

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
Autocorrelation function
Copula modeling
Probability distribution modeling

ABSTRACT

This study presents a method for using copulas to model the temporal variability of the clear-sky index, which in turn can be used to produce realistic time-series of photovoltaic power generation. The method utilizes the autocorrelation function of a clear-sky index time-series, and based on that a correlation matrix is set up for the dependency between clear-sky indices at N time-steps. With the use of this correlation matrix an N-dimensional copula function is configured so that correlated samples for these N time-steps can be obtained. Results from the copula model are compared with the original data set and an uncorrelated model based on zero correlated clear-sky index data in terms of distribution, autocorrelation, step changes and distribution. The copula model is shown to be superior to the uncorrelated model in these aspects. As a validation the model is tested with solar irradiance for two different geographical regions: Norrköping, Sweden and Hawaii, USA. The copula model is also applied to a set of bins of daily mean clear-sky index and the use of bins is shown to improve the results.

1. Introduction

The variability of solar irradiance on Earth's surface has an effect on many solar engineering applications, in particular those involving photovoltaic (PV) power generation (Lave et al., 2012; Bollen and Hassan, 2011; Kleissl, 2013; Widén, 2015). By quantifying the solar irradiance variability, it is possible to improve the design and operation of power systems where large amounts of distributed PV power might be injected into the grid (Bollen and Hassan, 2011; Widén, 2015; Lave et al., 2015). This is useful in order to avoid costly grid reinforcements (Bollen and Hassan, 2011; Holttinen et al., 2008). When high resolution data on solar irradiance is not available for certain time periods, or geographical locations, models that generate synthetic data can be useful (Bright et al., 2015). This paper extends the literature with a synthetic irradiance generator, based on a statistical method which has previously only been applied to spatial solar correlation modeling.

Solar irradiance, normalized via the clear-sky irradiance to the clear-sky index, has interesting variability on minute to instantaneous scale (Bright et al., 2015) and for such resolution the clear-sky index distribution can be modeled as a probability distribution with typically two or three peaks (Hollands and Huget, 1983; Hollands and Suehrcke, 2013; Munkhammar et al., 2015a; Widén et al., in press-a). Studies on clear-sky index ramp rates, and general temporal variability, include Markov-chain modeling (Bright et al., 2015), neural network modeling (Voyant et al., 2011) and pure sampling from probability distributions (Munkhammar et al., 2015a,b). Models vary in terms of input data complexity from utilizing cloud size, coverage and morphology (Bright et al., 2015; Smith et al., 2017) to analysis of the clear-sky index time-series (Munkhammar et al., 2015a,b; Munkhammar et al., 2017). A challenge is to not only obtain an accurate probability distribution, but to obtain a realistic synthetic time-series of the clear-sky index as well (Bright et al., 2017). These methods are intended to complement, and at best, improve on existing irradiation estimates from software, such as Meteonorm (2017), or irradiance from satellite data, see e.g. Engerer et al. (2017).

In terms of time-series realism, autocorrelation is a useful measure. Autocorrelation of the clear-sky index has been studied previously for instantaneous irradiance (Brinkworth, 1977; Skartveit and Olseth, 1992; Aguiar and Collares-Pereira, 1992; Hansen et al., 2010) and for step-changes as well (Hansen et al., 2010), 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 (Hansen et al., 2010; Perez et al., 2012). Models utilizing the autocorrelation of the clear-sky index include Brinkworth (1977), auto-regressive Gaussian (Aguiar and Collares-Pereira, 1992), neural networks (Voyant et al., 2011) and fractal cloud modeling (Lohmann et al., 2017). So-called clear-sky index generators are useful, since they use some existing data set of, e.g., lower resolution or averaged clear-sky index data to estimate higher resolution data

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

http://dx.doi.org/10.1016/j.solener.2017.09.028
Received 22 June 2017; Received in revised form 5 September 2017; Accepted 13 September 2017

(temporal or spatial), see e.g. Bright et al. (2015, 2017), Wegener et al. (2012), Ngoko et al. (2014), and Grantham et al. (2017). Autocorrelation functions of clear-sky index step changes have also been used in so-called virtual networks to study the variability of spatially distributed fleets of PV plants, see (Widén, 2015). This indicates that the autocorrelation function might be useful for both temporal and spatial solar irradiance studies.

Regarding spatial solar irradiance variability, copulas have been used to model the variability of solar irradiance in a spatial network (Munkhammar et al., 2017; Munkhammar and Widén, 2016). Copula modeling, being a multivariate statistical method for simulating correlated stochastic variables, is an increasingly common method for modeling correlation between stochastic variables (Nelsen, 2006). It has been used and described in studies of for example electric vehicle charging (Lojowska et al., 2012), wind power (Hagspiel et al., 2012), wave power (Li et al., 2016) and spatial solar irradiance variability (Munkhammar et al., 2017; Munkhammar and Widén, 2016).

This study develops an autocorrelation-based copula model for generating synthetic clear-sky index data, including the development of a model for generating synthetic clear-sky index data for binned daily mean clear-sky index levels. A copula-based model for quantifying the temporal variability of the clear-sky index on binned daily mean clear-sky index has not been done previously. The clear-sky index generator is, in similarity with for example (Skartveit and Olseth, 1992), based on utilizing only time-series as input. The model extends this type of modeling with more extensive autocorrelation statistics.

Daily clear-sky index bins have been used in other studies for generating synthetic clear-sky index data, in particular (Ngoko et al., 2014). In this paper the model simulations, and input data set, are based on minute resolution, where 120 min of each day for 365 days are used. General copula modeling and its application to temporal solar irradiance modeling is introduced in Section 2.2.

This study continues the work on correlation modeling of the clear-sky index in Munkhammar and Widén (2017), Munkhammar et al. (2017), and Munkhammar and Widén (2016), where the latter two studies presented models on spatial variability in the clear-sky index by using copulas and a spatiotemporal data set of the clear-sky index. If spatial modeling of solar irradiance is modeled under virtual network assumptions, generated from temporally shifted time-series based on cloud advection velocity, this model could also directly be used for spatial irradiance variability studies.

This paper is organized as follows. In Section 2 the methodology and data are presented, in Section 3 the results are presented and in Section 4 the results are discussed in a wider context.

2. Methodology

The methodology is organized so that the clear-sky index and an algorithmic description of the method is presented in Section 2.1, the copula modeling in Section 2.2, the goodness-of-fit statistic used in this study is presented in Section 2.3, the variability index is presented in Section 2.4 and the data set is presented in Section 2.5.

2.1. The clear-sky index

In order to focus on the temporal variability of instantaneous solar irradiance, the clear-sky index is used. 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:

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

The temporal copula model developed in this paper is based on the assumption that the clear-sky index is not dependent on time of day or season. This assumption can be contested, see for example (Smith et al., 2017). However, for expected time of day or seasonal variability in clear-sky index time series, bins of clear-sky index for time of day or season might be used instead.

Generally, throughout this study, it is assumed that the time-series of the clear-sky index is approximately *stationary*, a property of time-series which means that the joint probability distribution of a set of equivalently separated data points in a time-series does not change over time (Koopmans, 1995). If the clear-sky index is not expected to be stationary, the use of bins for specific time-periods, as mentioned, could perhaps improve this.

Since the clear-sky index is primarily defined only during daytime, when $G_c(t) > 0$, there is a maximum length of each daily vector of consecutive points of the clear-sky index. In order to have sufficient data for modeling this prompts the use of clear-sky index for a number of days (all days in a year in this study), each with the same length vector. Time, t, has resolution of minutes in this study.

Based on the assumption that the clear-sky index time-series for each day is stationary and a data set of clear-sky index for a number of days is available, the main model of the paper is based on the following algorithmic steps, where the copula modeling steps are described in detail in Section 2.2, see also (Munkhammar et al., 2017; Munkhammar and Widén, 2016; Nelsen, 2006) for more information.

1. Obtain a clear-sky index data set for N data points on M days, thus generating an $N \times M$ data set.
2. Estimate the autocorrelation function of the clear-sky index for each day, generating a data set of size $N \times M$.
3. Compute the mean autocorrelation function, for N steps, over all M days, generating a vector of the mean autocorrelation function of size N.
4. Define a correlation matrix, of size $N \times N$, from the N mean autocorrelation function values.
5. Define a copula based on the empirical distribution of the clear-sky index data set for N data points on all M days and the autocorrelation-based correlation matrix.
6. Generate, with the copula model, M number of synthetic correlated clear-sky index time-series with N data points each.

The concepts of these steps will be clarified, in a formal sense, in the following section. The proposed copula model is validated by comparing the synthetic copula-generated data sets with:

- Original clear-sky index data set of daily time-series. (The data set.)
- Generated uncorrelated daily data sets of clear-sky index time-series. (The uncorrelated model.)

The latter of these, the generated uncorrelated clear-sky index time-series, is the special case of the copula model with zero autocorrelation for non-zero lag, rendering the correlation matrix an identity matrix. This model is equivalent to a model of random sampling of the clear-sky index probability distribution and will generally be called the *uncorrelated model* in this study, since it lacks temporal correlation, which makes it equivalent to conventional probability distribution clear-sky index models (similar to the models used in e.g. Hollands and Suehrcke (2013) and Munkhammar et al. (2015a)). In spatial correlation modeling, the uncorrelated model was studied in for example (Munkhammar et al., 2017).

In addition to applying the copula model to mean autocorrelation function over all days (in step 3), a copula model is made for each bin of daily mean clear-sky index as well. The daily mean clear-sky index (calculated for the $N = 120$ data points for each day) was binned in five bins equally separated on the interval [0,1]. Then a mean autocorrelation function for the clear-sky index was obtained for all days within each bin. The algorithmic steps were then performed for each binned data set.

In terms of machine learning, it is customary to divide up training