Solar Energy 162 (2018) 500–509

Contents lists available at ScienceDirect

Solar Energy

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

Generating synthetic sequences of global horizontal irradiation

A.P. Grantham*, P.J. Pudney, J.W. Boland

University of South Australia, Mawson Lakes Boulevard, Mawson Lakes, SA 5095, Australia

A R T I C L E I N F O

2010 MSC:
00-01
99-00

Keywords:
Global horizontal irradiation
Solar
Synthetic sequences
Bootstrapping
Renewable energy

A B S T R A C T

When designing renewable energy systems it is common practice to use a short period of historical weather data or Typical Meteorological Year (TMY) data to evaluate the performance of a renewable energy system for a particular location. However, short periods of historical data or TMY data do not capture enough of the variation required to design a reliable renewable energy system. Longer data sequences are necessary to include a greater variety of sequences; synthetic sequences are useful because they can include sequences that have not occurred in the recorded data but are nonetheless as likely to occur as the observed data.

We propose a method to generate synthetic sequences of daily and hourly global horizontal irradiation (GHI) by developing a model to deal with the deterministic component of GHI, and then adding a stochastic component using a nonparametric bootstrapping technique. We use our synthetic daily and hourly models separately and reconcile them to match afterwards, unlike other studies that generate synthetic GHI data downscaled/interpolated from *observed* data. This is a fundamental difference to the literature.

The synthetic daily and hourly GHI sequences can be used, for example, as input for testing the performance and operation of a solar energy system for a wider range of scenarios than previously observed data. Further, one could incorporate synthetic GHI with other synthetic renewable energy data, such as synthetic wind farm electricity output. Both of these approaches would be useful when designing a reliable renewable energy system.

The synthetic sequences of daily and hourly GHI exhibit the same statistical properties as the real data. The two-sample Kolmogorov-Smirnov (KS) test shows that the distribution of the synthetic sequences of daily and hourly GHI match the distribution of the observed daily and hourly GHI respectively. The synthetic sequences of daily GHI have the same serial correlation structure as the observed data, an autoregressive model of order 1, AR(1), with similar AR(1) coefficients. Also, the synthetic sequences of hourly GHI have the same serial correlation structure as the observed data, an autoregressive model of order 3, AR(3), with similar AR(3) coefficients.

1. Introduction

1.1. Motivation

When designing renewable energy systems it is common practice to use historical weather data or Typical Meteorological Year (TMY) data (Marion and Urban, 1995; Wilcos and Marion, 2008) to evaluate the performance of a renewable energy system for a particular location. However, even twenty to thirty years of historical solar data, such as satellite solar data, isn't enough to capture enough of the variation required to design a reliable renewable energy system. Boland and Dik (2001, p.187) explains that "if the data set is not of sufficient length there will not necessarily be enough variation to provide estimates of performance under all possible conditions of the climate of the location".

This deficiency is also a problem when using TMY data, an alternative to historical data. Due to the way in which TMY data is constructed, the amount of variability that a particular climate variable, like global horizontal irradiation (GHI), may exhibit is underestimated. As a result, extreme weather sequences are filtered out. We demonstrate this in Section 7 where we show that our synthetic daily GHI data contains extreme sequences that do not occur in either the real data or TMY data. Consequently, TMY data is not suitable for designing reliable solar energy systems (Wilcos and Marion, 2008). Solar irradiation, for example, not only varies on small time scales, but also varies from year to year and hence one instance from TMY data is inadequate in representing a wide range of scenarios.

A sequence of values taken at regular time intervals is often called a time series. We use the term *sequence* to distinguish between a single sequence and multiple sequences. When modelling renewable energy systems, we are interested in how the system will behave with sequences of irradiation (for example) that may not appear in the

* Corresponding author.
E-mail address: adrian.grantham@mymail.unisa.edu.au (A.P. Grantham).

https://doi.org/10.1016/j.solener.2018.01.044
Received 17 November 2017; Received in revised form 8 January 2018; Accepted 15 January 2018
0038-092X/ © 2018 Elsevier Ltd. All rights reserved.

historical data or TMY data. Longer data sequences are necessary to include a greater variety of sequences; synthetic sequences, as suggested by Boland and Dik (2001), are useful because they can include sequences that have not yet occurred but may occur in the future.

1.2. Review of downscaling approaches

The term *synthetic* has a number of meanings in the literature to what we mean by the term in this paper. Many probabilistic forecasting studies use it to mean the construction of a full density forecast for a fixed time period ahead. A full density forecast provides information about all expected values, compared to a point forecast that provides information for only one expected value.

A number of studies in the literature on synthetic generation of GHI use the term when downscaling/interpolating between two observed values. Larrañeta et al. (2015) generate synthetic ten-minute data from observed hourly mean values—an improved method from Polo et al. (2011) who do the same downscaling. Peruchena et al. (2018) generate synthetic one-minute data from observed daily values. Hontoria et al. (2002) generate synthetic hourly data from observed daily data and note the synthetic hourly data could be generated from daily generation methods, like the method presented in this paper. Other downscaling studies include (Grantham et al., 2013, 2017; Fernández-Peruchena et al., 2015).

We use the terms *synthetic sequences* or *synthetic data* to mean scenarios over all time. Our synthetic sequences describe a realistic sequence of estimates of GHI, where each subsequence has not necessarily occurred in the data set from which the synthetic data is generated. The subsequences are equally as likely to occur as those that have occurred in the past or will occur in the future. For example, when we generate a synthetic daily GHI value for January 1, the value has the same statistical properties as observed values for the same time of year. However, while our synthetic daily value is for January 1, it is not from a particular year. A fundamental difference between what we do in this paper and those studies described above is that we generate synthetic daily data and synthetic hourly data independently and then reconcile them to match. Whereas, other studies do some synthetic—they go from observed to higher frequency synthetic data.

Other terminology like *scenario generation* is also used in the literature. While scenario generation may sound like synthetic sequences, it does not mean the same thing in the context of this paper. For example, Osório et al. (2015) refer to scenario generation as the generation of representative values of hourly wind power generation, that are equally as likely to occur as the observed wind power generation data; they then use these representative values to estimate forecasting errors in solving the day-ahead unit commitment (UC) problem.

1.3. Review of synthetic modelling approaches

Studies on the generation of synthetic sequences, by our definition, are very limited. Synthetic estimation of rainfall on various time scales is proposed in Piantadosi et al. (2008). Magnano et al. (2008) generate synthetic sequences of daily and half-hourly temperature. The synthetic half-hourly data is used to generate half-hourly synthetic sequences of electricity demand in South Australia (Boland, 2010; Magnano and Boland, 2007).

Rastogi and Andersen (2015) borrow from Magnano and Boland (2007) to create synthetic weather data from TMY data (temperature only) for use in building simulations. Because the synthetic data is created from TMY data, their approach cannot accurately reflect years that are significantly different to the base TMY data.

Aguiar et al. (1999) assesses the value of TMY data built from observed and synthetic data for use in building thermal simulations. The study compares the thermal performance of a building using TMY data built from observed data (the classical method), synthetic weather data, and synthetic TMY data. The synthetic weather data and synthetic TMY data are built using the CLIMED software (developed by Aguiar) for the Mediterranean climate. CLIMED uses a top-down approach to synthetically generate monthly, then daily, and finally hourly synthetic data with GHI being the master variable with other climate variables generated with reference to it. The algorithms for the monthly, daily and hourly synthetic generation use clearness index to deal with the seasonality and Markov chains, either ARMA models or a Markov Transition Matrix, to deal with a serial correlation structure.

Fernández-Peruchena et al. (2015) propose a Multiscale Stochastic (MUS) methodology for generating high-frequency synthetic solar data (or other meteorological variables). The MUS methodology is a multistep process that brings together other well-known methodologies to generate annual and monthly GHI and direct normal irradiation (DNI) totals which are then downscaled to daily, then hourly and then subhourly GHI and DNI. While synthetic annual totals and coherent monthly totals of GHI and DNI are generated, Fernández-Peruchena et al. (2015) don't generate higher frequency data. Instead, a review of methodologies for downscaling monthly to one-minute is provided. In this context, *coherent* means the synthetic monthly totals for a calendar year sum to the synthetic annual total.

Morf (2013) generates sequences of instantaneous global horizontal irradiance with a time scale in the magnitude of seconds. Their model uses the probability of sunshine or not and cloudiness models and then they check the performance of their method by comparing the characteristics of the daily simulated to the observed daily data. The major difference between this paper and what we do is that we use time series data to generate synthetic sequences and they do not.

Nielsen et al. (2017) generate synthetic yearly DNI values using six different approaches which then are used to calculate probability of exceedance (PoE) values. PoE values are used by financing and banking institutions to determine the feasibility of solar energy systems.

Ngoko et al. (2014) use Markov chains trained on one-minute observed data. This Markov chain is geographically dependent. Bright et al. (2015) generate one-minute clear or cloudy sky indicators using Markov chains. During clear periods they use a clear-sky model, during cloudy periods they use a clear-sky index value extracted from distributions that were discovered from a study of the relationship between clear-sky index and observations of okta (the amount of cloud cover). Their work is extended in Bright et al. (2017) to a spatio-temporal method. Their methods requires observational data of okta, wind speed, cloud base height and atmospheric pressure. We only require GHI. These studies use clearness index (Ngoko et al., 2014) and clear-sky index (Bright et al., 2015, 2017) methods to deal with seasonality whereas we use Fourier series.

Boland (2008) (and related studies (Boland, 1997; Boland and Dik, 2001)) uses classical time series modelling techniques to generate synthetic daily and hourly GHI time series. Classical time series analysis decomposes a time series into a deterministic component (trend/seasonal/correlation) and a random component. Daily and hourly models are generated to fit the daily and hourly observed GHI data respectively. The daily synthetic generation process follows the daily fitted model and includes a white noise term (variation) which is sampled from a beta distribution. A similar process is also used for the hourly synthetic generation process but the white noise term is sampled from a normal distribution. However, a parametric approach imposes assumptions on the underlying distribution structure of the white noise component.

Our work presented in this paper extends the work by Boland (2008). As in Boland (2008), we use classical time series techniques to model daily GHI but in this paper, the white noise term (variation) is obtained through *nonparametric bootstrapping*. Nonparametric bootstrapping is a more robust method for estimating white noise than the parametric approach used in Boland (2008) because it is less sensitive to extreme values of the noise. We use a nonparametric approach because it does not impose any parametric assumptions on the underlying distribution structure of the white noise component.

Download English Version:

https://daneshyari.com/en/article/7935751

Download Persian Version:

https://daneshyari.com/article/7935751

Daneshyari.com