Energy and Buildings 82 (2014) 580–591

Contents lists available at ScienceDirect

Energy and Buildings

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

Restoration of long-term missing gaps in solar radiation

Oluwaseyi T. Ogunsola, Li Song *

University of Oklahoma, Norman, OK, United States

ARTICLE INFO

Article history:
Received 19 January 2014
Received in revised form 28 July 2014
Accepted 30 July 2014

Keywords:
Solar radiation
Weather data
Singular spectrum analysis
Missing data restoration
Clear sky model
Temperature-based approach

ABSTRACT

Solar radiation is an important climatic variable and widely used in building performance monitoring and analysis. However, due to sensor malfunction, data transmission problems, and quality assurance issues, there are often short-term or long-term missing data on solar radiation. These gaps are challenging for engineers involved in building performance monitoring and control. This paper examines and compares three different approaches, namely, singular spectrum analysis (SSA), statistically adjusted solar radiation (SASR), and the temperature-based approach (TBA), for restoring missing solar data. The TBA, SASR, and SSA are applied to fill artificial gaps which are generated continuously as representative of up to 25 days missing data in actual hourly solar radiation data of Oklahoma City North (OKCN) for 2012 and Albuquerque for 2005. Results show that SSA outperforms the other methods for filling solar radiation gaps of up to 5 days. For gaps up to 20 days, the SSA and TBA have similar performance. For gaps larger than 20 days, the TBA is more suitable. The SASR performs similarly to the TBA for dry and sunny Alburquerque climates, but worse in OKCN. Accuracy of the SSA decreases with increasing gap lengths. The study concludes by recommending appropriate methods for different gap lengths.

1. Introduction

Solar radiation is an important variable in building performance simulation, monitoring and analysis. Measured data, such as solar radiation, has shown to be useful in improving the accuracy and reliability of building energy simulations especially for commercial buildings due to the existence of large glazed area [1,2]. Solar radiation is made up of several broad classes of electromagnetic radiation. It is one of the climatic variables recorded by most weather stations. Solar radiation data can have significant impact on energy use, as well as building performance, control, monitoring, and analysis. It is a very important variable in the design and operation of heating, ventilation and air-conditioning (HVAC) systems. The availability of reliable solar radiation data is very critical for the selection of appropriate control measures or for maximization of energy savings opportunities. Although solar radiation data is recorded by weather stations all over the world, a certain level of measured data is often not available. Data from nearby stations may not be a true representation of the solar radiation at neighboring locations. Also, the available data may contain missing data for several days. This might be caused by sensor error and damage, time and cost limitations, data transmission problems, or data quality assurance methods. Therefore, there are often short or long-term missing solar radiation data. These issues present challenges for building designers, modelers, operators, and other users of solar radiation data, since the missing gaps have to be filled before the data can be utilized. Unreasonable interpolation of missing solar radiation data can lead to costly errors, particularly when applied to design, or assessment of control efforts in building systems operations.

Many methods have been investigated for filling gaps in time-series data, and these methods are not necessarily limited to solar data restoration. Chen and Claridge [3] evaluated Single Variable Regression, Polynomial models, Lagrange Interpolation, and Linear Interpolation models to estimate values for 1–6 h gaps in weather, heating, and cooling measurements for commercial buildings. Hu et al. [4,5] also utilized Lagrange Interpolation, Linear Interpolation, and Spatial–temporal Interpolation to fill short- and long-term missing gaps in climate variables such as solar radiation, temperature, and wind speed. Hocke and Kämpfer [6] filled data gaps using the Lomb-Scargle periodogram, which is well-known for estimation of the power spectral density of sampled data. The Lomb-Scargle periodogram provides least squares spectrum in case of uneven data, but reduces to the Fourier transform when there are evenly sampled data. In the Lomb-Scargle method, the time series is usually modified in a frequency domain by imposing a confidence limit on the amplitude or by considering only those spectral components within a specified frequency domain. The time series data

* Corresponding author. Tel.: +1 405 325 1714.
E-mail address: lsong@ou.edu (L. Song).

http://dx.doi.org/10.1016/j.enbuild.2014.07.088
0378-7788/

Nomenclature

CV_MAE	coefficient of variation of the mean absolute error (%)
CV_ME	coefficient of variation of the mean error (%)
CV_RMSE	coefficient of variation of the root mean square error (%)
D	total daily radiation (MJ m^{-2})
G	extra terrestrial radiation (MJ m^{-2})
H	hourly radiation (MJ m^{-2})
MAE	mean absolute error (MJ m^{-2})
ME	mean error (MJ m^{-2})
N	total length of data (days)
RMSE	root mean square error (MJ m^{-2})
T	dry bulb temperature (°C)
W	windows length for SSA (days)
Z	elevation (m)
d	day of year
n	number of data points
y	actual solar radiation (MJ m^{-2})
ACSM	ASHRAE clear sky model
OKCN	Oklahoma City North
SASR	statistically adjusted solar radiation
SSA	singular spectrum analysis
SVD	singular value decomposition
TBA	temperature based approach

Greek symbols

δ	angle of declination (°)
φ	latitude of the location (°)
ω	hour angle (°)
λ	Eigenvalue
θ	hour from noon

Subscripts

s	sunset
i	i-th data
n	day of year
oh	horizontal
c	corrected
r	ratio
max	maximum
min	minimum

Superscripts

est	estimated
T	transpose

would then be reconstructed by taking the inverse Fourier transform of the relevant spectral components. The method is deemed capable of reproducing steep data gradients. Sjögren et al. [7] evaluated the energy performance of buildings based on incomplete monthly data using the energy signature approach, which compensates for missing temperature and household electricity data by assuming different consumption profiles. The method is seen as one of the effective ways to get around missing data in building total energy use. Kondrashov and Ghil [8] proposed an approach to fill gaps in time series based on Singular Spectrum Analysis (SSA). SSA incorporates elements from a wide range of mathematical fields including classical time series analysis, multivariate statistics and geometry, dynamical systems, as well as signal processing [9]. It describes the structure of a time series using a simpler elementary series which determines the trend, oscillations, and noise. The SSA technique performs well in the simultaneous extraction of harmonics and trend components.

SSA has been widely applied for estimation, trending, and forecasting in control systems studies and mathematics. The term *singular spectrum* comes from the spectral (Eigenvalue) decomposition of a matrix A into its set (spectrum) of Eigenvalues. The origins of SSA can be traced to Broomhead and King [10,11] and Broomhead et al. [12] in the application to chaos theory. More properly, SSA should be called the analysis of time series using the singular spectrum [13]. SSA aims to decompose an original time series into the sum of independent components. These components represent the trend, oscillations, and noise, which make up a typical time series. Using adjacent data, it is possible to recover the trend and oscillations, which could be used for forecasting, or gap filling, depending on the application. Unlike the Fourier approach and other function-based approaches which generate a set of components directly from the time series under study, SSA does not rely on a priori-defined function [14]. It has been widely used for forecasting and recovery of missing data in time series. The possible applications of SSA include mathematics, economics, finance, market research, meteorology, oceanography, social sciences, market research, and any field with apparently complex series with a potential structure [15,9]. To the best of our knowledge, there are few or no studies demonstrating the application of the SSA method in recovering missing data in solar radiation.

There have been several studies on estimating missing solar radiation data. ASHRAE RP 1413 [16] developed the statistically adjusted solar radiation (SASR) method for filling climatic data gaps for use in building performance monitoring and analysis. In the SASR method, correction was made to the ASHRAE clear sky model 2005 [17] which tended to over-predict the solar radiation on non-clear sky days. The correction was made based on the difference between maximum and minimum dry-bulb temperature for the day. Dimas et al. [18] estimated missing data in hourly solar radiation based on atmospheric transmittance and available meteorological data. The accuracy was deemed to be strongly linked to the accurate determination of beam atmospheric transmittance, which was assigned based on humidity, and by finding correlation between relative humidity, clearness index, and beam transmittance. Hokoi et al. [19] developed a statistical method to generate synthetic weather data as an alternative to the use of measured hourly data. The method of estimation was based on the Kalman filter, and an autoregressive moving average was developed for solar radiation time series. Solar radiation has also been obtained from other climatic variables. For example, Taesler and Andersson [20] computed solar radiation successively using routine meteorological observations of humidity, cloudiness, visibility and snow cover while taking into account different orientation, inclination, and screening from surrounding terrain, buildings or other objects. Similarly, the air temperature-based approach (TBA) uses maximum and minimum air temperature to estimate atmospheric transmissivity (e.g., [21,22]). A range of methods is available for conversion of sunshine duration to daily global solar radiation values (e.g., [23,24]). However, the physical mechanism used in observation of sunshine hours can result in large measurement errors. In general, most of the existing solar data restoration methods need other variables. For example, TBA and SASR are both temperature-based models, and their accuracies are heavily dependent on reliable dry-bulb temperature measurements. In order to eliminate this limitation, the SSA method is explored in this paper as an alternative to recover missing data in solar radiation.

Therefore, the aim of this study is the development and validation of standard procedures for filling missing gaps in solar radiation, to enable complete data set creation. This study also