Renewable Energy 103 (2017) 70–80

Contents lists available at ScienceDirect

Renewable Energy

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

A guideline to select an estimation model of daily global solar radiation between geostatistical interpolation and stochastic simulation approaches

D.I. Jeong [a, *], A. St-Hilaire [b], Y. Gratton [b], C. Bélanger [b], C. Saad [c]

[a] Centre ESCER (Étude et Simulation du Climat à l'Échelle Régionale), Université du Québec à Montréal, Montreal, Canada
[b] INRS-ETE, University of Québec, Québec, Canada
[c] Environment and Climate Change Canada, Montreal, Canada

ARTICLE INFO

Article history:
Received 29 June 2016
Received in revised form
16 October 2016
Accepted 12 November 2016
Available online 15 November 2016

Keywords:
Artificial neural networks
Geostatistical interpolation
Global solar radiation
Spatial correlation
Temperature
Relative humidity

ABSTRACT

This study compares geostatistical interpolation and stochastic simulation approaches for the estimation of daily global solar radiation (GSR) on a horizontal surface in order to fill in missing values and to extend short record length of a meteorological station. A guideline to select an approach is suggested based on this comparison. Three geostatistical interpolation models are developed using the nearest neighbor (NN), inverse distance weighted (IDW), and ordinary kriging (OK) schemes. Three stochastic simulation models are also developed using the artificial neural network (ANN) method with daily temperature (ANN(T)), relative humidity (ANN(H)), and both (ANN(TH)) variables as predictors. The six models are compared at 13 meteorological stations located across southern Quebec, Canada. The three geostatistical interpolation models yield better performances at stations located in a high density area of GSR measuring stations compared to the three stochastic simulation models. The guideline suggests an optimal approach by comparing a threshold distance, estimated according to a performance criteria of a stochastic simulation model, to the distance between a target and its nearest neighboring station. Additionally, the spatial correlation strength of daily GSRs and the at-site correlation strength between daily GSRs and the predictor variables should be considered.

1. Introduction

Global solar radiation (GSR) on a horizontal surface of the earth is an important variable for many analyses involving agricultural and plant growth, air and water temperatures, environmental and biological risk, and solar electric generation. However, instruments measuring solar irradiation (i.e., Kipp or Eppley pyranometers) are relatively expensive and difficult to manage [1], compared to those of common meteorological variables such as air temperature, precipitation, and relative humidity. Therefore, meteorological stations for GSR are generally less abundant than those for the common meteorological variables. Furthermore, observed GSR datasets are usually short timeseries and have large gaps of missing values.

Geostatistical interpolation approaches can be adopted to fill in missing values and to extend short record length of the GSR at a station using observed GSR data on the other stations located near the desired station. Kriging [2–7], nearest neighbor [4], and inverse distance weighted average [5,8,9] approaches have been applied frequently for the spatial interpolation.

At-site physical and statistical approaches can also be used for GSR simulations. Physical models (e.g. [10–12]) use complex physical interactions between the GSR and the terrestrial atmosphere, such as the Rayleigh scattering, radiative absorption by ozone and water vapour and aerosol extinction. Stochastic simulation models (e.g. [13–20]) use empirical relationships between GSR and meteorological covariables such as sunshine hours, temperature, and relative humidity at a desired station. This study considers stochastic simulation models as they are relatively simple to develop and require fewer input variables compared to physical models [16,17]. Although linear and non-linear regressions as well as artificial neural networks (ANNs) can be employed to drive empirical relationships between the common meteorological variables and the GSR, many studies [16–18,20–22] have shown the

* Corresponding author. Centre ESCER, Université du Québec à Montréal, 201 Ave. President-Kennedy, Montreal, Quebec H3C 3P8, Canada.
E-mail address: jeong@sca.uqam.ca (D.I. Jeong).

http://dx.doi.org/10.1016/j.renene.2016.11.022
0960-1481/

superiority of ANN approaches to regression-based approaches.

Sunshine duration, one of the most explanatory variables for GSR simulation [18,21,23], has not been recorded at most meteorological stations in Canada since 1999 due to its difficulty of measurement [1]. Temperature [13–15,17–19,24–28] and relative humidity [18,21] are alternative covariables although they have weaker correlation with GSR compared to sunshine duration [18].

Geostatistical interpolation and statistical simulation approaches for GSR estimation have been applied separately in many studies, however, they have been rarely compared in an application study. Therefore, this study compares geostatistical interpolation and statistical simulation approaches to fill in missing values and to extend short record length of daily GSR timeseries. The spatial interpolation approaches considered include the nearest neighbor, the inverse distance weighted, and the ordinary kriging methods. The stochastic simulation models include three ANN-based models with daily temperature and/or daily relative humidity as input variables. The six models are applied at 13 meteorological stations located across southern Quebec (45.1–50.3 °N and 64.2–79.0 °W), Canada. Furthermore, a guideline to choose an approach between the geostatistical interpolation and the statistical simulation approaches is provided for the estimation of daily GSR on the study area.

2. Methodologies

2.1. Geostatistical interpolation models

Three geostatistical interpolation models are developed based on nearest neighbor (NN), inverse distance weighted (IDW), and ordinary kriging (OK) schemes for daily GSR. The NN model employs the simplest algorithm among the three models. This model selects the value of the nearest station to the location of interest and does not consider the other values of neighboring stations in order to yield a piecewise-constant interpolation map.

The IDW interpolation algorithm adopts the assumption that the interpolation value at a location of interest is inversely proportional to the distances of nearby stations. The interpolation value of the model is a weighted average of the values of multiple stations and the weight assigned to each nearby station diminishes as the distance from the interpolation point to that station increases. The IDW model interpolates the daily GSR value $R(x_0)$ at an ungauged location x_0 from observations $R(x_i)$ at locations $x_1,\ldots,x_n$ as follows:

$$\widehat{R}(x_0) = \sum_{i=1}^{n} w_i R(x_i) \tag{1}$$

$$w_i = \frac{1/d_i}{\sum_{j=1}^{n} 1/d_j},\ i = 1,\ 2,\ \ldots,\ n \tag{2}$$

where $\widehat{R}(x_0)$ is an interpolated value of $R(x_0)$ and d_i represents distance between $R(x_0)$ and $R(x_i)$.

Kriging is a geostatistical interpolation technique based on the linear least square estimation algorithm. Ordinary kriging (OK) is the most common among many kriging approaches. OK estimates the best linear unbiased estimator based on a linear model. The interpolation value of the OK at a location x_0 is given by the following equation:

$$\widehat{R}(x_0) = \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix}^{\prime T} \begin{pmatrix} R(x_1) \\ \vdots \\ R(x_n) \end{pmatrix} \tag{3}$$

where $w_1,\ldots,w_n$ are the weights of the OK that fulfill the unbiased condition $\sum_{i=1}^{n} w_i = 1$. The weights are obtained by the below OK equation system:

$$\begin{pmatrix} w_1 \\ \vdots \\ w_n \\ \mu \end{pmatrix} = \begin{pmatrix} \gamma(x_1,x_1) & \cdots & \gamma(x_1,x_n) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma(x_n,x_1) & \ldots & \gamma(x_n,x_n) & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix}^{-1} \begin{pmatrix} \gamma(x_1,x_0) \\ \vdots \\ \gamma(x_n,x_0) \\ 1 \end{pmatrix} \tag{4}$$

where $\mu = E[R(x)]$ is a Lagrange parameter employed to minimize the kriging error under the unbias condition, which is assumed to be an unknown constant in the OK. $\gamma(x_i, x_j)$ is a variogram function to calculate the spatial dependency between $R(x_i)$ and $R(x_j)$. Several variogram functions are available such as exponential, Gaussian, and spherical models. In this study, the spherical variogram function is selected based on trial and error examination. The variogram is estimated for each day, based on observed daily GSR dataset of nearby stations. The detail descriptions of variogram models and ordinary kriging can be found in Refs. [29,30].

To verify the interpolation performances of the three models, a leave-one-out cross-validation approach is employed. Among the observations at n stations, GSR values of one of those stations are interpolated using the observations at the remaining n-1 stations. This process is repeated for all the observation stations. The interpolated $\widehat{R}(x_i)$ is then compared to the associated observation $R(x_i)$ at each station in order to evaluate the performance of the interpolation models.

2.2. Stochastic simulation models

Three stochastic simulation models are developed to estimate daily GSR using the ANN approach as a transfer function and daily maximum and minimum temperatures and/or daily mean relative humidity as input variables. Feed forward ANNs have been frequently employed to simulate GSR [16–18,20,21] from the meteorological input variables. This study also employs a three-layer feed forward ANN model, which includes an input layer, a single hidden layer, and an output layer of computation nodes. The ANN models are trained by the Bayesian regularization back-propagation (BRBP) algorithm, which is a network training function that updates the weight and bias values according to the Levenberg-Marquardt optimization [31]. An important issue in ANN modelling is the determination of the number of hidden nodes. Fletcher and Goss [32] suggested that the optimal number of hidden nodes could be within $(2p^{0.5} + o) \sim (2p+1)$, where p and o are the numbers of independent and dependent variables, respectively. The hyperbolic tangent sigmoid function is employed for the hidden layer and the linear function is used for the output layer. Detailed descriptions of these various activation functions are provided in Ref. [31]. The three ANN models used to simulate daily GSR series from daily meteorological variables are as follows:

$$\widehat{R} = ANN(T_{\max}, T_{\min}, R_a) \tag{5}$$

$$\widehat{R} = ANN(H, R_a) \tag{6}$$

$$\widehat{R} = ANN(T_{\max}, T_{\min}, H, R_a) \tag{7}$$

where $T_{\max}$ and $T_{\min}$ are daily maximum and minimum temperatures (°K) and H is daily mean relative humidity in a given day. The ANN represents the three-layer feed forward ANN trained by the BRBP algorithm. The R_a is the solar irradiation on a horizontal surface at the top of the atmosphere, which is a function of latitude and Julian day of a site. It is calculated by using the standard