Renewable and Sustainable Energy Reviews 94 (2018) 732–747

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

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

Evaluating the effect of air pollution on global and diffuse solar radiation prediction using support vector machine modeling based on sunshine duration and air temperature

Junliang Fan[a], Lifeng Wu[a,b,*], Fucang Zhang[a,c], Huanjie Cai[a,c], Xiukang Wang[d], Xianghui Lu[b], Youzhen Xiang[a]

[a] *Key Laboratory of Agricultural Soil and Water Engineering in Arid and Semiarid Areas of Ministry of Education, Northwest A&F University, Yangling 712100, China*
[b] *School of Hydraulic and Ecological Engineering, Nanchang Institute of Technology, Nanchang 330099, China*
[c] *Institute of Water-saving Agriculture in Arid Areas of China, Northwest A&F University, Yangling 712100, China*
[d] *College of Life Sciences, Yan'an University, Yan'an 716000, China*

ARTICLE INFO

Keywords:
Global dimming
Air pollutant
Air quality index
Suspended particulate matters
Ozone

ABSTRACT

Increasing air pollutants attenuate surface solar radiation, and thus can be influential variables for solar radiation prediction. In this study, six air pollutants of $PM_{2.5}$, PM_{10}, SO_2, NO_2, CO and O_3 as well as air quality index (AQI) were chosen for analyzing their single and integrated effects on daily global and diffuse solar radiation (R_s and R_d) prediction. Seven single air pollution parameters, 15 combinations of two parameters and 20 combinations of three parameters were considered using Support Vector Machine (SVM) based on sunshine duration or air temperature. Daily meteorological and air pollution data between January 2014 and December 2015 from China's capital city of Beijing were used to train SVM models and data from January 2016 to December 2016 for testing. Results show that AQI was the most relevant air pollution parameter for both R_s and R_d prediction, followed by O_3 for R_s and by $PM_{2.5}$ for R_d with slight difference as that of AQI. The combination of PM_{10} and O_3 and the combination of $PM_{2.5}$ and O_3 were the most influential combination of two air pollution inputs for R_s and R_d prediction, respectively. The combination of $PM_{2.5}$, PM_{10} and O_3 was the most optimal combination of three air pollution inputs for both daily R_s and R_d prediction. Compared with SVM models without considering air pollution, the accuracy of SVM models with the most influential combinations of one, two and three air pollution inputs was improved by 13.9%, 19.8% and 22.2% in terms of RMSE for sunshine-based R_s, respectively. The corresponding values were 15.2%, 22.0% and 22.8% for temperature-based R_s, 16.1%, 21.5% and 24.5% for sunshine-based R_d, and 16.8%, 22.0% and 23.3% for temperature-based R_d. The results demonstrate the importance of appropriate selection of air pollution inputs to improve the accuracy of R_s and R_d prediction in air-polluted regions.

1. Introduction

Among the renewable and sustainable energy resources (e.g. solar, wind, biomass, geothermal and hydroelectric), solar energy has attracted much attention due to its abundant availability on the Earth's surface and being environmentally-friendly [57,31,45]. Global solar radiation (R_s) and its component of diffuse solar radiation (R_d) at a given location are of great importance for agricultural and hydrological modeling as well as the optimal design and application of solar energy systems [2,18,25,29,38,41]. Nevertheless, unlike other meteorological variables (e.g. sunshine duration and air temperature), reliable measurements of global and diffuse solar radiation are not available at many worldwide locations, particularly in developing countries, due to the high costs and the difficulty of installation and maintenance of measuring instruments like Pyranometers [23]. Taking China as an example, there are 726 long-term weather stations, where 98 stations record global solar radiation, but only 17 stations measure diffuse solar radiation [26]. Therefore, various techniques have been proposed to predict R_s and R_d, e.g. empirical models [1,4,8,24,15] and machine learning models [11,16,17,34,37,40].

Abbreviations: ANFIS, Adaptive Neuro Fuzzy Inference System; ANN, Artificial Neural Networks; API, Air Pollution Index; AQI, Air Quality Index; BTH, Beijing–Tianjin–Hebei; GDP, Gross Domestic Product; MAE, Mean absolute error ($MJ\,m^{-2}\,d^{-1}$); R^2, Coefficient of determination; RF, Random Forest; RMSE, Root mean square error ($MJ\,m^{-2}\,d^{-1}$); SVM, Support Vector Machine

* Corresponding author at: School of Hydraulic and Ecological Engineering, Nanchang Institute of Technology, Nanchang 330099, China.
E-mail address: china.sw@163.com (L. Wu).

https://doi.org/10.1016/j.rser.2018.06.029
Received 28 February 2018; Received in revised form 13 June 2018; Accepted 13 June 2018
1364-0321/

Nomenclature

C	Penalty parameter of the error
CO	Carbon monoxide (mg m^{-3})
k	Number of observations
n	Sunshine duration (h)
N	Maximum sunshine duration (h)
NO_2	Nitrogen dioxide (μg m^{-3})
O_3	Ozone (μg m^{-3})
$PM_{2.5}$	Suspended particulate matters < 2.5 μm in aerodynamic diameter (μg m^{-3})
PM_{10}	Suspended particulate matters < 10 μm in aerodynamic diameter (μg m^{-3})
R_a	Extra-terrestrial solar radiation (MJ m^{-2} d^{-1})
R_d	Diffuse solar radiation (MJ m^{-2} d^{-1})
R_s	Global solar radiation (MJ m^{-2} d^{-1})
SO_2	Sulfur dioxide (μg m^{-3})
T_{max}	Maximum temperature (°C)
T_{min}	Minimum temperature (°C)
φ	Higher-dimensional feature space
ω	Weights vector
ε	Tube size
λ	Regularization parameter
γ	Minimum loss
Ω	Regularization term

The empirical models are most commonly used due to their model simplicity and low computational costs [22,23,44]. Over the past few decades, many efforts have been made to predict R_s and R_d from various types of empirical models, e.g. sunshine-based models [24,25,36,5,6] and temperature-based models [16,17,22,30,50]. Generally, the sunshine-based empirical models provide better estimates than those based on maximum/minimum temperature [13,54]. However, when lack of sunshine duration measurements, the temperature-based empirical models are highly preferred for solar radiation prediction due to air temperature being the most available meteorological variable at any stations around the world [22]. Although empirical models have been widely employed for solar radiation prediction, they are difficult to deal with nonlinear and multidimensional relationships in noisy environments [28]. Thus, various machine learning techniques have been applied to predict solar radiation, e.g. Artificial Neural Networks (ANN) ([7,51,32]), Support Vector Machines (SVM) [37,40] and Adaptive Neuro Fuzzy Inference System (ANFIS) [34,56], etc. Among these machine learning models, the SVM model has been recently employed for predicting solar radiation from sunshine duration or air temperature data owing to its higher accurate predictions compared with the other models [21,33,39,47].

Apart from sunshine duration and air temperature, other meteorological and geographical factors like precipitation, relative humidity and location (longitude and latitude), were also used for solar radiation

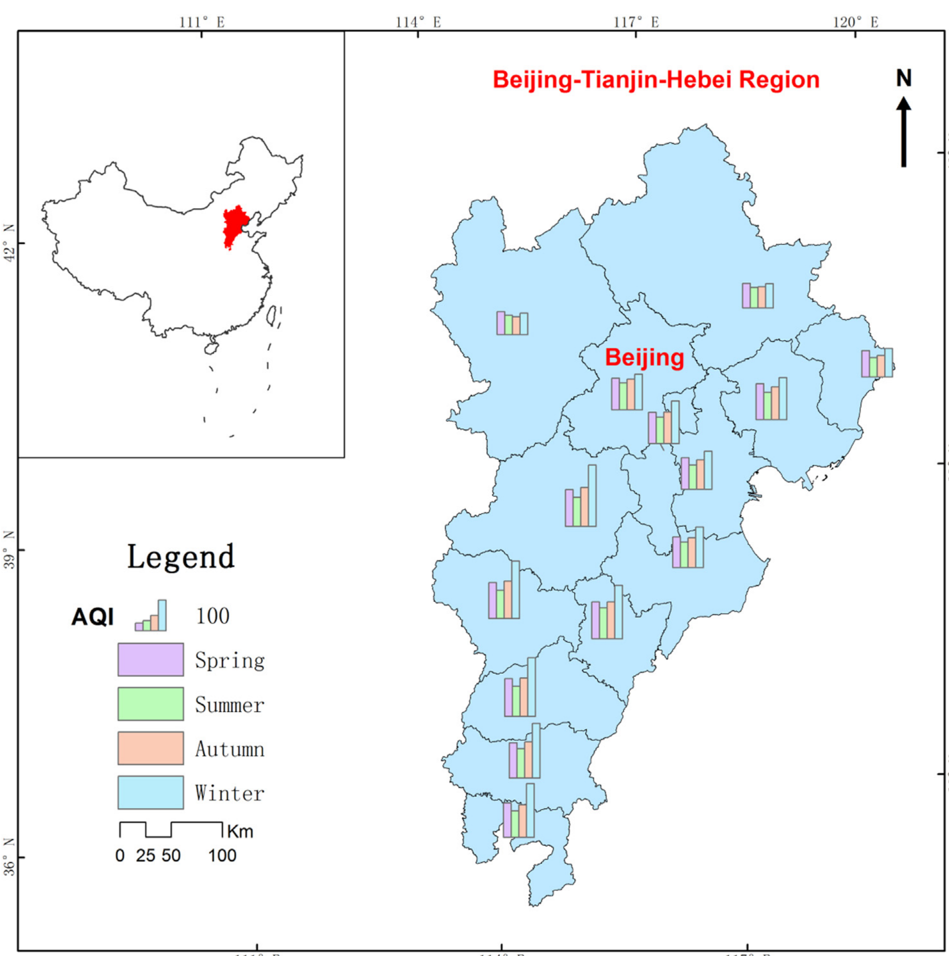

Fig. 1. The geographical locations of the 14 major cities in the Beijing-Tianjin-Hebei (BTH) region of China. The average seasonal AQI during 2014–2016 in each city is also presented.