Renewable Energy 128 (2018) 155–162

Contents lists available at ScienceDirect

Renewable Energy

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

# A support vector machine approach to estimate global solar radiation with the influence of fog and haze

Wanxiang Yao [a, b], Chunxiao Zhang [c, *], Haodong Hao [d], Xiao Wang [e], Xianli Li [c]

[a] Tianjin Key Laboratory of Civil Structure Protection and Reinforcement, Tianjin Chengjian University, Tianjin, 300384, PR China
[b] State Key Lab of Subtropical Building Science, South China University of Technology, Guangzhou, 510640, PR China
[c] School of Energy and Safety Engineering, Tianjin Chengjian University, Tianjin, 300384, PR China
[d] School of Control and Mechanical Engineering, Tianjin Chengjian University, Tianjin, 300384, PR China
[e] School of Economics and Management, Tianjin Chengjian University, Tianjin, 300384, PR China

ARTICLE INFO

*Article history:*
Received 7 February 2018
Received in revised form
25 April 2018
Accepted 19 May 2018
Available online 21 May 2018

*Keywords:*
Support vector machines
Global solar radiation
Fog and haze
Air quality index

ABSTRACT

In recent years, fog and haze occurred frequently, due to energy crisis and environmental pollution. Fog and haze have significant scattering-weakening effect on solar radiation, resulting in a severe weaken to solar radiation received on a horizontal surface. In this paper, air quality index (AQI) is taken as an additional input parameter, and some new models for estimating global solar radiation on a horizontal surface are proposed based on a support vector machine (SVM). The accuracy of SVM-1 and SVM-2 models are compared and analyzed, and the results show that the performance of SVM-2 models with an extra input parameter AQI are generally improved, for which the R value is promoted from 0.848 to 0.876, the NSE value is lifted from 0.682 to 0.740, the RMSE value is reduced from 0.114 to 0.102, and the MAPE value is decreased from 9.257 to 8.214. Comparing with existing models, SVM models proposed in this paper can improve the accuracy of global solar radiation models. If AQI is used as an additional input parameter to estimate global solar radiation, the accuracy will be further improved.



## 1. Introduction

In recent years, population growth has led to the increasingly intensified energy crisis and environmental pollution. All countries have focused their research on renewable energy. As a renewable energy source with no pollution, large reserves and wide distribution, solar energy has been widely used [1–3]. Solar radiation data plays an important role in solar energy conversion, building energy conservation, and regional resource assessment. However, in some developing countries, due to lack of observation equipment, equipment maintenance or other reasons, solar radiation data is seriously missing. Therefore, it is particularly necessary to establish daily global solar radiation models.

Before reaching the Earth's surface, solar radiation is absorbed and scattered by the atmosphere. Due to the influence of weather conditions, atmospheric conditions and other factors, solar radiation received on the earth's surface shows a random variation trend. Some scholars have studied the influence of conventional meteorological parameters (e.g. latitude, sunshine duration, temperature, relative humidity, etc.) on a horizontal surface of solar radiation, and many empirical models have been proposed based on a large number of historical data [4–6]. In addition, some researchers have tried to estimate global solar radiation on a horizontal surface by coupling satellite images with ground observation data [7–9], which is more suitable to estimate solar radiation in a large scale (larger than 10 km), especially for special weather conditions (such as overcast days, cloudy days, rain and snow days, etc.). Solar radiation is influenced by many factors at the same time. Considering the relative simplification of functional relationship, empirical models can only reflect the influence of one or several parameters on solar radiation. However, the interaction of multiple parameters is rarely considered.

With the development of artificial intelligence and bio-anthropophic technology, many researchers apply neural networks to solar energy, especially in the study of solar radiation [10–12]. Because neural network is not restricted by conventional functions and can consider the interaction between multiple parameters at the same time, it has unique advantages to estimate solar radiation. Support vector machine (SVM) has a wide range of applicability compared to other neural networks, which can not

* Corresponding author.
*E-mail address:* zcx_tj@126.com (C. Zhang).

https://doi.org/10.1016/j.renene.2018.05.069
0960-1481/

**Nomenclature**

| | |
|---|---|
| $T_{max}$ | Daily maximum air temperature (°C) |
| $T_{min}$ | Daily minimum air temperature (°C) |
| $\Delta T_{air}$ | Daily air temperature difference (°C) |
| $\Delta T_{sur}$ | Daily surface temperature difference (°C) |
| $N$ | Number of days between 1 (January 1st) and 365 or 366 (December 31st) |
| $S$ | Sunshine duration (h) |
| $S_0$ | Daily maximum possible sunshine duration (h) |
| $WS$ | Wind speed (m/s) |
| $RH$ | Relative humidity (%) |
| $PM2.5$ | Particulate matter on a scale of 2.5 |
| $PM10$ | Particulate matter on a scale of 10 |
| $AQI$ | Air quality index (dimensionless) |
| $SVM$ | Support vector machine |
| $ANN$ | Artificial neural network |
| $GP$ | Genetic programming |
| $R$ | Correlation coefficient (dimensionless) |
| $NSE$ | Nash-Sutcliffe Equation (dimensionless) |
| $RMSE$ | Root mean square error (dimensionless) |
| $MAPE$ | Mean absolute percentage error (%) |

only deal with linear problems, but also map nonlinear problems into high-dimensional linear space by kernel function to ensure the adaptability of this algorithm. In addition, based on statistical learning theory, SVM uses the principle of structural risk minimization to learn from finite samples, and seeks trade-off between approximation accuracy and approximation function complexity of given data, so that this algorithm has strong generalization ability. Therefore, lots of solar radiation models have been established by various scholars based on SVM algorithm in the past decades [13–18]. Chen et al. [13] proposed a series of models which use a combination of seven temperatures as input parameters, and linear, polynomial and radial basis functions as kernel functions. These models use support vector machine to estimate monthly average solar radiation. Results show that the model which uses polynomial as kernel function and $T_{max}$ and $T_{min}$ as input parameters is the best. Based on historical data (solar radiation, cloud cover, relative humidity and wind speed, etc.), Zeng et al. [14] used SVM algorithm to predict atmospheric transmissivity and then converted it into solar energy according to latitude and the time of day. Through verification with the National Solar Radiation Database, this model shows a good consistency. If additional meteorological variables are used, especially cloud cover, the accuracy of this model will be further improved. Chen et al. [16] established global solar radiation models based on sunshine duration using support vector machine learning, and compared SVM models of seven different input parameters with five empirical models. The results demonstrate that the performance of SVM model is better than empirical models. Meanwhile, combining sunshine duration with other meteorological parameters can improve the accuracy of models. Z. Ramedani et al. [15] proposed a support vector regression method to predict global solar radiation. Two SVR models are studied, where one is radial basis function model (SVR-rbf) and the other is a polynomial function model (SVR-poly). The estimated performance of SVR-rbf model is better than SVR-poly model, which can effectively improve the accuracy and shorten calculation time. Olatomiwa et al. [17] used sunshine duration, maximum and minimum temperature as input parameters, and combined support vector machine with firefly algorithm (SVM-FFA) to predict monthly average solar radiation on a horizontal surface. Compared artificial neural network (ANN) and Genetic Programming (GP) models, the SVM-FFA models have higher prediction accuracy. S. Shamshirband et al. [18] combined support vector machine with wavelet analysis to establish a coupled model for estimating solar radiation, and tested with data of Iran (cloud cover and clearness index). As a result, the model has a good agreement with measured data.

In addition, environmental pressure is increasing with the process of industrialization. There are different degrees of fog and haze in many regions of China, in which Beijing-Tianjin-Hebei region is particularly serious [19,20]. The number of air pollution days in Beijing was 143 days in 2017, accounting for 39.2% of the whole year, with severe and serious pollution for 27 days. On sunny days, solar radiation is absorbed and scattered by the upper atmosphere; when it is overcast or cloudy days, especially heavy fog and hazy weather, solar radiation is not only passing through atmosphere, but also low-level clouds and particles in air near the ground. Solar radiation is scattered, reflected and refracted, and it is severely weakened when reaching the ground. Therefore, near-surface atmosphere seriously affects the acceptable solar radiation on a horizontal surface. The near-surface atmospheric conditions can be characterized by air quality index (AQI), which is almost directly proportional to the concentration of particulate matter (PM) in the air [21,22]. Some researchers proposed AQI as an additional input parameter to establish solar radiation models. For example, Yao et al. [23] established empirical model of daily diffuse solar radiation based on the data of nearly 55 years in Beijing, China, and modified it with AQI. Result demonstrates that the accuracy of this model is improved. Based on meteorological data of Tehran in Iran for one year, Masoud Vakili et al. [24] established daily global solar radiation models by using an artificial neural network which input parameters include temperature, relative humidity, wind speed and air particulate matter (PM).

The main purpose of this paper is using SVM algorithm to realize the learning mechanism of finite samples based on the principle of minimizing structural risk, and investigate the influence of different input parameters on output performance of these models, especially the contribution of AQI to output performance. In this paper, based on the influence of fog and haze, SVM neural network is applied to establish daily global solar radiation models of Beijing. The influence of different input parameters (surface temperature difference, air temperature difference, relative humidity, sunshine duration and AQI) on the accuracy of SVM models are compared and analyzed. In addition, these new SVM models and existing models are compared and analyzed based on statistical parameters. The results of this paper can provide a reference for estimating solar radiation in the regions which are lacking of detection equipment, and provides a basis for the evaluation, conversion, and utilization of solar energy resources, especially for regions where environmental pollution is more serious.

## 2. Method and data

### 2.1. Support vector machine

Support vector machines (SVM) were originally proposed by V. Vapnik [25,26] for classification, non-linear regression and other related fields. The theoretical basis of SVM is statistical learning theory, more precisely, SVM is an approximate realization of structural risk minimization. According to statistical learning theory, the actual risk of regression estimation is related to empirical risk and confidence range. The empirical risk is related to training sample, and the confidence range is related to Vapnik-Chervonenkis dimension (VC dimension) of learning machine and number of training samples. When these two factors are both small, that is, not only empirical risk is small, but also VC dimension is