Energy Conversion and Management 91 (2015) 433–441

Contents lists available at ScienceDirect

Energy Conversion and Management

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

Support vector regression based prediction of global solar radiation on a horizontal surface

Kasra Mohammadi [a,*], Shahaboddin Shamshirband [b,*], Mohammad Hossein Anisi [b], Khubaib Amjad Alam [c], Dalibor Petković [d]

[a] Faculty of Mechanical Engineering, University of Kashan, Kashan, Iran
[b] Department of Computer System & Technology, Faculty of Computer Science & Information Technology, University of Malaya, 50603 Kuala Lumpur, Malaysia
[c] Department of Software Engineering, Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur, Malaysia
[d] University of Niš, Faculty of Mechanical Engineering, Deparment for Mechatronics and Control, Aleksandra Medvedeva 14, 18000 Niš, Serbia

ARTICLE INFO

Article history:
Received 27 September 2014
Accepted 4 December 2014
Available online 30 December 2014

Keywords:
Support vector regression methodology
Global solar radiation estimation
Sunshine hour
Maximum possible sunshine hour
Empirical models

ABSTRACT

In this paper, the support vector regression (SVR) methodology was adopted to estimate the horizontal global solar radiation (HGSR) based upon sunshine hours (n) and maximum possible sunshine hours (N) as input parameters. The capability of two SVRs of radial basis function (rbf) and polynomial basis function (poly) was investigated and compared with the conventional sunshine duration-based empirical models. For this purpose, long-term measured data for a city situated in sunny part of Iran was utilized. Exploration was performed on both daily and monthly mean scales to accomplish a more complete analysis. Through a statistical comparative study, using 6 well-known statistical parameters, the results proved the superiority of developed SVR models over the empirical models. Also, SVR-rbf outperformed the SVR-poly in terms of accuracy. For SVR-rbf model on daily estimation, the mean absolute percentage error, mean absolute bias error, root mean square error, relative root mean square error and coefficient of determination were 10.4466%, 1.2524 MJ/m^2, 2.0046 MJ/m^2, 9.0343% and 0.9133, respectively. Also, on monthly mean estimation the values were 1.4078%, 0.2845 MJ/m^2, 0.45044 MJ/m^2, 2.2576% and 0.9949, respectively. The achieved results conclusively demonstrated that the SVR-rbf is highly qualified for HGSR estimation using n and N.

1. Introduction

Recently, renewable energy sources such as solar energy are harnessed globally to mitigate some negative environmental issues including climate change due to large-scale fossil fuel exploitation. Solar energy is regarded as the most attractive renewable energy source owing to its broad availability and feasibility of exploitation in many regions across the globe [1,2]. As such, many technologies have been conceived for the sake of solar energy utilization, particularly in sunny regions of the world. However, precise knowledge of solar radiation on different time scales is of great significance for the successful simulation, operation and monitoring of solar energy technologies [3–6]. Due to the unavailability or even lack of reliable measured solar radiation data in some regions, especially in developing countries, solar radiation estimation based on precise models and techniques has always played a notable role. Over the past decades, a vast number of empirical models have been developed to estimate the global solar radiation using several meteorological and geographical parameters. Among all elements used as inputs, sunshine hours, broadly measured at meteorological stations, have often been endorsed as one of most important parameters for the accurate estimation of global solar radiation.

The first and extensively utilized model proposed for estimation of the global solar radiation on a horizontal surface is that of Angström [7]. The proposed model correlated the ratio of monthly mean daily horizontal global solar radiation to its corresponding value in a clear day with the ratio of monthly mean daily relative sunshine duration. However, the fundamental difficulty was lack of accessibility to correct value of global solar radiation in a clear day. Therefore, Prescott [8] recommended utilizing the extraterrestrial horizontal global solar radiation instead of clear day solar radiation. The achieved model globally known as Angström–Prescott model is considered as the simplest sunshine duration-based empirical model. Subsequently, in efforts to improve the estimation exactness, a series of modifications have been performed

* Corresponding authors. Tel.: +98 9121955642 (K. Mohammadi), tel.: +60 146266763 (S. Shamshirband).

E-mail addresses: kasra278@yahoo.com (K. Mohammadi), shamshirband@um.edu.my (S. Shamshirband).

http://dx.doi.org/10.1016/j.enconman.2014.12.015
0196-8904/

mainly by increasing the order of relative sunshine duration [9–12] or using various functional forms [13–15].

Due to significance of precise solar radiation data for solar experts, architects, agriculturists and hydrologists, applications of different artificial intelligence techniques have gained more importance in the realm of horizontal global solar radiation estimation.

In recent years, the support vector machines (SVMs) have been employed as a robust technique in a variety of scientific and engineering applications [16–24]. SVMs are a soft computing method that has obtained importance in environmental issues. Basically, there are two fundamental classes of support vector machines: (*a*) support vector classification (SVC) and (*b*) support vector regression (SVR). SVC is a learning framework utilizing a high-dimensional peculiarity space [25–28], while SVR focused around measurable learning hypothesis and structural risk minimization rule and has been effectively utilized for nonlinear frameworks [29,30]. The correctness of an SVM model is to a great extent relies on determination of its model parameters. Nonetheless, inspection of the literature reveals that despite the growing applications of SVMs in various areas, only a few studies have been carried out to use this technique for estimation of horizontal global solar radiation [31–35].

Considering the lack of enough investigation on application of SVR technique in solar radiation area and also importance of sunshine hours for estimation of global solar radiation, the chief aim of this research work is to appraise the viability of two SVRs of radial basis function (rbf) and polynomial basis function (poly) for estimation of horizontal global solar radiation using only two inputs parameters of sunshine hours (n) and maximum possible sunshine hours (N). As a case study, long-term measured databases for city of Isfahan located in central desert part of Iran have been used. The merit of developed SVR models is evaluated in terms of both daily and monthly basis to draw more conclusive conclusion. To ensure the suitability level of developed SVR models their accuracies are compared with four well-known sunshine duration-based empirical models established in this study for Isfahan and also a previously published model. To offer a thorough comparative study, the comparisons and performances validation are carried out based on several widely utilized statistical indicators.

2. Case study and data description

The city of Isfahan, capital of Isfahan province of Iran, has been considered as a sole case study. Isfahan city (Fig. 1) is situated in a sunny belt part of the country with geographical location of 32°37′N and 51°40′E. Isfahan enjoys mild and dry weather condition with low level of participation throughout the year, which is only around 123 mm [36]. The monthly average temperature varies from 3.4 °C to 28.9 °C and the yearly average is 16.2 °C. The monthly average relative humidity varies between 25% and 60% with the annual average of 40%. On the basis of the Köppen classification, the arid desert cold climate of Isfahan is categorized as BWk [37]. As the city is located in sunny region of Iran, it enjoys significant solar energy potential during the entire year. For Isfahan, according to the long-term averaged data, the daily global solar radiation on a horizontal surface throughout the year is in the range of 8.98–28.94 MJ/m^2 with the yearly average of 19.47 MJ/m^2. Also, the monthly mean daily global solar radiation on a horizontal surface varies in the range of 9.3029.34 MJ/m^2 throughout the year with the yearly average of 19.69 MJ/m^2.

In this study, the long-term measured daily global solar radiation on a horizontal surface (H) and sunshine hours (n) provided by Iranian Meteorological Organization (IMO) for Isfahan for the period of 1985–1991 and 1998–2003 were utilized.

The precision of the models developed to estimate the solar radiation is chiefly influenced by the quality of raw data utilized. The data cleaning procedure generally aims at enhancing the data quality by checking and filtering them from any uncertainty or erroneous. In horizontal global solar radiation data used in this study, there were some missing and also unreliable values possibly due to instruments' malfunction. For the period of 1992–1997 which were not included in this study, the horizontal global solar radiation contained either so many missing or unreliable values. To overcome this issue and enhance the quality of raw data, the following procedure was applied in present study:

(1) To identify the incorrect global solar radiation values, the daily clearness index (K_T) was computed and the values which were out of range of $0.015 < K_T < 1$ were eliminated [38,39].
(2) A month containing more than 5 days missing or unreliable global solar radiation values was completely extracted from the databases. Additionally, for a month with less than 5 days the missing or inaccurate values were substituted by proper values obtained using interpolation [38,39].
(3) To achieve further accuracy, for the omitted months the sunshine duration data were also excluded from the data series.

It is worth mentioning that clearness index (K_T) is the ratio between global solar radiation incidents on a horizontal surface (H) to extraterrestrial horizontal solar radiation on a horizontal surface (H_o). H_o in any geographical location is a constant value for each specific day, irrespective of change of year. However, solar attenuation occurs as radiation passes through the atmosphere due to some atmospheric phenomenon such as aerosol extinction, cloud extinction, Rayleigh scattering and so on. Therefore, in the available solar radiation data all values of H should be smaller than H_o, which means $K_T < 1$.

The available data for this study are divided into two parts of training and testing data sets. 9 years data sets for the period of 1985–1991 and 1998–1999 are utilized for training. While the remaining 4 years data sets for the period of 2000–2003 are used for testing. In fact, on daily basis estimation the system is trained using 3285 days and tested by 1460 days. Also, for monthly mean daily basis calculation 108 months data is used for training and 48 months data is applied for testing.

3. Methodology

3.1. Support vector regression application

The main principle of SVMs is to do the data correlation through non-linear mapping. If a way of computing the inner product in a feature space is available directly as a function to the original input points, it is feasible to build a non-linear learning machine, which is known as a direct computation method of a kernel function, denoted by K.

The flexibility of the SVM is in regard to the kernel functions that implicitly convert the data to a higher-dimensional feature space. Results in the higher-dimensional feature space correspond to the results of the original, lower-dimensional input space. There are some methods that employ non-linear kernels for regression problems. One kernel function is the radial basis function. The main benefit of the radial basis function is computationally efficiency since radial basis function training needs only the solution of a set of linear equations instead of the lengthy and computationally demanding quadratic programming problem. Therefore, the radial basis function with parameter σ is adopted in this study. The non-linear radial basis kernel function is defined as:

$$K(x, x_i) = \exp\left(-\frac{1}{\sigma^2} x - x_i^2\right) \quad (1)$$