Theor Appl Climatol (2014) 115:627–638
DOI 10.1007/s00704-013-0924-y

ORIGINAL PAPER

Evaluation of support vector machine for estimation of solar radiation from measured meteorological variables

Ji-Long Chen · Guo-Sheng Li

Received: 16 April 2012 / Accepted: 6 May 2013 / Published online: 22 May 2013

Abstract Solar radiation is an essential and important variable to many models. However, it is measured at a very limited number of meteorological stations in the world. Developing method for accurate estimation of solar radiation from measured meteorological variables has been a focus and challenging task. This paper presents the method of solar radiation estimation using support vector machine (SVM). The main objective of this work is to examine the feasibility of SVM and explore its potential in solar radiation estimation. A total of 20 SVM models using different combinations of sunshine ratio, maximum and minimum air temperature, relative humidity, and atmospheric water vapor pressure as input attributes are explored using meteorological data at 15 stations in China. These models significantly outperform the empirical models with an average 14 % higher accuracy. When sunshine duration data are available, model SVM2 using sunshine ratio and air temperature range is proposed. It significantly outperforms the empirical models with an average 26 % higher accuracy. When sunshine duration data are not available, model SVM19 using maximum temperature, minimum temperature and atmospheric water vapor pressure is proposed. It significantly outperforms the temperature-based empirical models with an average of 18 % higher accuracy. The remarkable improvement indicates that the SVM method would be a promising alternative over traditional approaches for estimation of solar radiation at any locations.

J.-L. Chen
Chongqing Institute of Green and Intelligent Technology, CAS, Chongqing 401122, China

J.-L. Chen · G.-S. Li (✉)
Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, China
e-mail: ligscas@163.com

1 Introduction

A good knowledge of solar radiation is essential for many applications, including agricultural, ecological, hydrological and soil–vegetation–atmosphere transfer models (Liu et al. 2009b). Despite its significance, accurate long-term records of solar radiation are not widely available due to the cost of measuring equipment and its difficult maintenance and calibration (Hunt et al. 1998). Lack of sufficient radiation data has been reported in many countries such as the USA (Ball et al. 2004; Garcia and Hoogenboom 2005), United Kingdom (Rivington et al. 2005), Egypt (Mossad 2005), India (Polo et al. 2011) and China (Chen and Li 2012a). This has led researchers to develop methods to estimate solar radiation for the sites where no direct solar radiation data are available.

A number of alternatives have been developed to estimated solar radiation. Major groups of those methods include stochastic algorithm (Richardson 1981; Hansen 1999; Wilks and Wilby 1999), empirical relationship (Ångström 1924; Prescott 1940; Hargreaves et al. 1985), spatial interpolation (Ducco et al. 1998; Bechini et al. 2000; Xia et al. 2000; Soltani et al. 2003) and satellite-based models (Pinker et al. 1995; Schillings et al. 2004; Janjai et al. 2009; Şenkal 2010). The stochastic method may be useful for generating average theoretical scenarios. However, the generated data cannot be used for model calibration and validation since it neither generates extreme weather condition, nor produces data which match actual weather condition at a particular time of interest (Wallis and Griffiths 1995; Liu and Scott 2001). Satellite can provide data continuously in space and time, thus the satellite-based model is promising for parameterization of solar radiation over large, remote area. However, the low sampling frequency and coarse spatial resolution of satellite-based methods renders them inadequate for site-specific application (Pinker et al. 1995). Many satellite-based models have difficulties to produce accurate results, the deviations between the results and the measurements are up to 40 % (Vignola et al. 2007; Janjai et al. 2009). Spatial interpolation technique can predict values at unknown locations and create surface from surrounding points. The main problem, however, is the lack of sufficient stations of solar radiation measurement.

The empirical relationship method using measured meteorological variables is attractive because of the good data availability. In particular, the well-known sunshine- and temperature-based models are widely used. Ångström (1924) first proposed a simple linear equation to estimate solar radiation using sunshine duration which was further modified by Prescott (1940) (A-P). Several modified versions of the A-P model have subsequently been made by adding more additional available meteorological variables such as air temperature (Chen et al. 2004), relative humidity (Ojosu and Komolafe 1987; Gopinathan 1988; Ododo et al. 1995) and atmospheric water vapor pressure (Garg and Garg 1982; Abdalla 1994). However, the accuracy improvements of those modified versions were not validated in many works (Kuye and Jagtap 1994; Ertekin and Yaldiz 1999; Mossad 2005; Wu et al. 2007; Chen and Li. 2012b). Gueymard et al. (1995) posed some fundamental questions that how can the simple A-P model be improved to generate more accurate long-term solar radiation data by those available climatological variables, and believed these research questions remain unanswered. He also illustrated and stressed the need for new avenues of approach that could lead to improved models.

Although the sunshine-based method is generally more accurate (Iziomon and Mayer 2002; Podestá et al. 2004; Trnka et al. 2005), it is often limited since sunshine duration is not commonly measured as air temperature. In this context, Hargreaves et al. (1985) (H-S) proposed an equation to estimate solar radiation using air temperature. Many revised formulations were subsequently made by adding more additional meteorological variables (De and Stewart 1993; Hunt et al. 1998; Thornton and Running 1999; Chen et al. 2006). Although some authors claimed that their revised models outperformed the H-S model, this may not always be the cases in many comparative studies (Liu and Scott 2001; Manual et al. 2003; Ball et al. 2004; Wu et al. 2007; Chen and Li 2012b). Moreover, models using air temperature sometimes cannot produce the results that meet the requirement of applications (Choisnel et al. 1992; Chen et al. 2004; Zhou et al. 2005). So, how can the temperature-based models be improved to generate more accurate solar radiation data? To fully take advantage of the easy availability of commonly measured meteorological variables, it is very important and urgent to explore new approaches that can produce more accurate long-term solar radiation data than the traditional empirical methods which seem to have difficulty in increasing the estimation accuracy.

This is perhaps more critical now than ever before, because of the worldwide interest in renewable energy utilization in areas where measured solar radiation data are scarce, and the increasing needs of accurate long-term solar radiation data for studying the global climate change (Gueymard et al. 1995). In recent years, as the development of computational technology and sophisticated statistical methods, machine learning method has been increasingly studied and shown as a powerful tool in forecasting and regression. Among the developed machine learning methods, support vector machine (SVM), originally developed by Vapnik (1995), has been widely applied in computer, environment and hydrology researches (Lee and Verri 2003; Lu and Wang 2005; Tirusew et al. 2006; Stanislaw and Konrad 2007). A number of studies have proved that SVM shows better performance than neural network and other traditional statistical models (Dibike et al. 2001; Liong and Sivapragasam 2002; Kecman 2005; Chen et al. 2011). Despite successes in many fields, there is no literature on the application of SVM in solar radiation estimation. Thus, the main objectives of this study are (1) to examine the feasibility of SVM and explore its potential in estimation of solar radiation from measured meteorological variables, including sunshine duration, maximum and minimum air temperature, relative humidity and atmospheric water vapor pressure; (2) to investigate the effects of different inputs on accuracy of SVM; and (3) to propose a selection strategy on optimal SVM models for estimation of solar radiation under different situations. Two scenarios are considered: (1) all of those meteorological data are available, and (2) only sunshine duration data are not available.

2 Materials and methods

2.1 Study sites and data collection

A total of 15 stations with long-term available records of solar radiation across China were used in this work. The mapping of stations roughly range from 18° to 46°N (latitude), from 79° to 127°E (longitude), and from 6 to 3,648 m altitude, covering five major climate zones (Fig. 1): cool temperature zone (stations Haerbing, Hailiutu and Wulumuqi), warm temperate zone (Beijing, Xian, Jiuquan and Hetian), subtropical zone (Shanghai, Changsha, Chengdu, Kunming and Guangzhou), tropical (Sanya) zone and Qinghai–Tibet plateau climate zone (Geermu and lasha). Table 1 shows the detailed geographical information of these meteorological stations.

Monthly mean daily solar radiation (MJ m^{-2}), sunshine duration (h), air temperature (°C) including maximum and minimum, relative humidity (%) and atmospheric water vapor pressure (kPa) were used in this study. Solar radiation

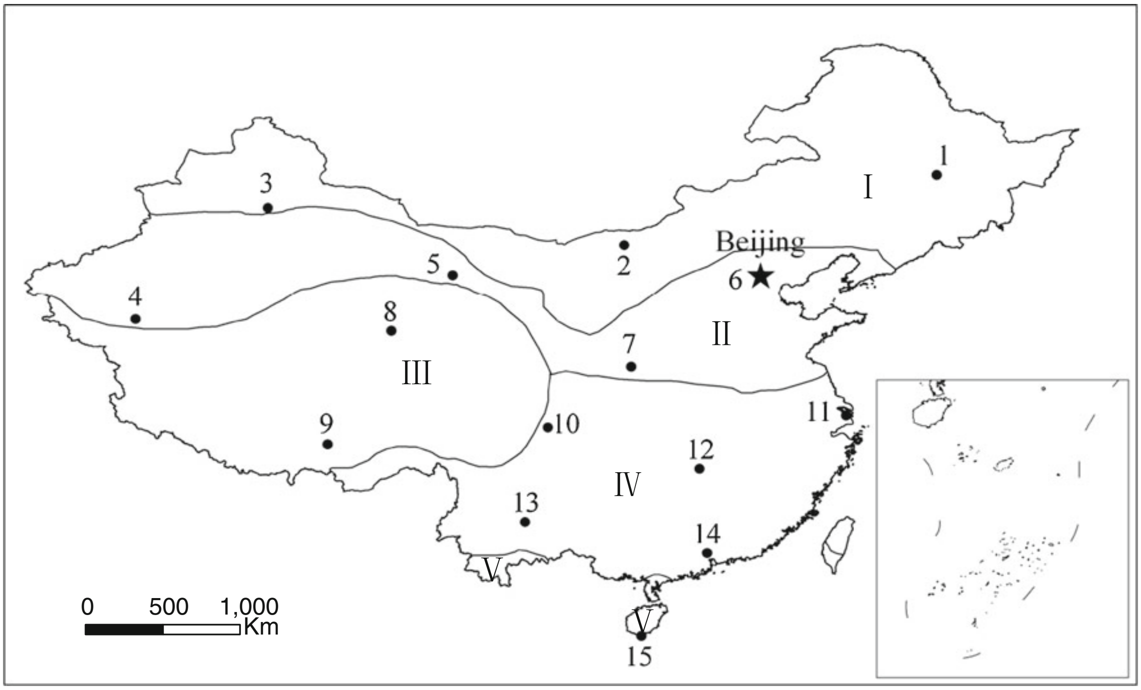

Fig. 1 Location of the study meteorological stations (stations are numbered in compliance with Table 1). I, II, III, IV, and V represent cool temperature, warm temperate, Qinghai–Tibet plateau climate, subtropical, and tropical zones, respectively

is measured using pyranometer (DFY-4) (5 %). The type of pyranometer used by the China Meteorological Bureau was change since 1993. However, the homogeneity of the radiation series is probably little affected, as the pyranometer models have been calibrated to the same standard (Yang et al. 2009) following the guide of the WMO (Liu et al. 2009a; Chen et al. 2011). Other relevant meteorological variables are routinely measured following the guidelines of the WMO. The data were obtained from the National Meteorological Information Center (NMIC), China Meteorological Administration (CMA). The period of records ranges from 18 to 41 years covering the period between 1970 and 2010. Quality control tests were conducted by the supplier. Two data sets were created for each station. About 75 % of the total records were used for training SVM models and calibrating the empirical models, and the remainder for validation. The detailed temporal periods are listed in Table 1.

Table 1 Detailed information of the study meteorological stations

Station ID	Station name	Latitude (N)	Longitude (E)	Altitude (m)	Training period	Validation period
1	Haerbing	126.77	45.75	142.3	1970–2000	2001–2010
2	Haliutu	108.52	41.57	1,288.0	1992–2005	2006–2010
3	Wulumuqi	87.65	43.78	935.0	1970–2000	2001–2010
4	Hetian	79.93	37.13	1,374.5	1970–2000	2001–2010
5	Jiuquan	98.48	39.77	1,477.2	1993–2005	2006–2010
6	Beijing	116.47	39.81	31.3	1970–2000	2001–2010
7	Xian	108.93	34.31	397.5	1970–1996	1997–2005
8	Geermu	94.91	36.42	2,807.6	1970–2000	2001–2010
9	Lasha	91.13	29.67	3,648.7	1970–2000	2001–2010
10	Chendu	104.02	30.67	506.1	1970–1994	1995–2003
11	Shanghai	121.48	31.41	6.0	1991–2005	2006–2010
12	Changsha	112.92	28.22	68.0	1987–2004	2005–2010
13	Kunming	102.68	25.02	1,892.4	1970–2000	2001–2010
14	Guangzhou	113.33	23.17	41.0	1970–2000	2001–2010
15	Sanya	109.52	18.23	5.9	1992–2005	2006–2010

2.2 Theory of support vector machine

A brief theory of SVM is presented in this section. More detailed information on the subject can be found in the work of Vapnik (1995, 1998). The SVM, originally developed by Vapnik and his coworkers, has been widely applied in classification, regression and forecasting. Comparing to traditional learning machine methods, there are several attractive characteristics of SVM. First of all, SVM implements the principle of Structural Risk Minimization, which attempts to minimize an upper bound of generalization error rather than minimize the local training error. This is the significant difference from commonly used principle of empirical risk minimization, which is employed by the traditional learning machine methods. Secondly, SVM estimates the regression using a set of kernel functions which are defined in a high dimensional space. Thirdly, SVM uses a risk function consisting of the empirical error and a regularization term which is derived from the structure risk minimization principle.

Given a set of data points $G=\{(x_i,d_i)\}_i^n$ ($\boldsymbol{x}_i$ is the input vector, d_i is the desired value and n is the total number of data patterns), SVM approximates the function using the following form:

$$f(x) = w\phi(x) + b \tag{1}$$

where ϕ (x) is the high-dimensional feature space which is nonlinearly mapped from the input space x. The coefficients w and b are estimated by minimizing the regularized risk function below (Vapnik 1995, 1998):

$$R_{\mathrm{SVMs}}(C) = C\frac{1}{n}\sum_{i=1}^{n} L(d_i, y_i) + \frac{1}{2}\|w\|^2 \tag{2}$$

In the regularized risk function, the term $C\frac{1}{n}\sum_{i=1}^{n} L(d_i, y_i)$ is the empirical error (risk), and measured by function $L\varepsilon$ given below:

$$L\varepsilon(d,y) = \begin{cases} |d-y|-\varepsilon |d-y|\geq\varepsilon \\ 0 \quad \text{otherwise} \end{cases} \tag{3}$$

The term $\frac{1}{2}\|w\|^2$ is the regularization term. C is referred to as the regularized constant and determines the trade-off between the empirical risk and the regularization term. Increasing the value of C will result in the relative importance of the empirical risk with respect to the regularization term to grow. ε denotes the tube size, and it is equivalent to the approximation accuracy placed on the training data points.

To obtain the estimations of w and b, Eq. 2 is transformed to the primal function given by Eq. 3 by introducing the positive slack variables ζ_i and ζ_i* as follows (Vapnik 1995):

$$\text{Minimize } R_{\mathrm{SVMs}}\left(w, \zeta^{(*)}\right) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\zeta_i + \zeta_i^{*}) \tag{4}$$

$$\text{Subjected to } \begin{array}{c} d_i - w\phi(x_i) - b_i \leq \varepsilon + \zeta_i \\ w\phi(x_i) + b_i - d_i \leq \varepsilon + \zeta_i^{*}, \zeta_i^{*} \geq 0 \end{array} \tag{5}$$

Finally, by introducing Lagrange multipliers and exploiting the optimality constraints, the decision function given by Eq. 1 has the following explicit form:

$$f(x, a_i, a_i^{*}) = \sum_{i=1}^{n}(a_i - a_i^{*})K(x, x_i) + b \tag{6}$$

The detail computation procedure can be found in the work of Vapnik et al. (1996).

The term $K(x_i,x_j)$ is called kernel function, the value of kernel function $K(x_i,x_j)$ is equal to the inner product of two vectors $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ in the feature space $\varphi(x_i)$ and $\varphi(x_j)$, that is, $K(x_i,x_j)=\varphi(xi)*\varphi(xj)$. Thus, all computations related to $\varphi(x)$ can be performed by the kernel function in input space. The elegance of using the kernel function is that one can deal with feature spaces of arbitrary dimensionality without having to explicitly compute the map $\varphi(x)$. Any function satisfying Mercer's condition can be used as kernel function (Vapnik et al. 1996). In this study, the common radial basis

Table 2 The study SVM models with different input attributes

Model	Input attributes[a]	Model	Input attributes[a]
SVM1	S/S_o	SVM11	S/S_o, T_{max}, T_{min}, VP
SVM2	S/S_o, T_{max}–T_{min}	SVM12	S/S_o, T_{max}, T_{min}, RH, VP
SVM3	S/S_o, RH	SVM13	T_{max}–T_{min}
SVM4	S/S_o, VP	SVM14	T_{max}–T_{min}, RH
SVM5	S/S_o, RH, VP	SVM15	T_{max}–T_{min}, VP
SVM6	S/S_o, T_{max}–T_{min}, RH	SVM16	T_{max}–T_{min}, RH, VP
SVM7	S/S_o, T_{max}–T_{min},VP	SVM17	T_{max}, T_{min}
SVM8	S/S_o, T_{max}–T_{min}, RH, VP	SVM18	T_{max}, T_{min}, RH
SVM9	S/S_o, T_{max}, T_{min}	SVM19	T_{max}, T_{min}, VP
SVM10	S/S_o, T_{max}, T_{min}, RH	SVM20	T_{max}, T_{min}, RH, VP

[a] S, S_o, T_{max}, T_{min}, RH, and VP denote monthly mean sunshine duration, potential sunshine duration, maximum air temperature, minimum air temperature, relative humidity, and atmospheric water vapor pressure, respectively

Table 3 Root mean square errors (RMSE) of sunshine duration based SVM models[a]

Station	SVM1	SVM2	SVM3	SVM4	SVM5	SVM6
Haerbing	0.998 (7.94 %)^{11}	0.913 (7.27 %)^{7}	0.989 (7.87 %)^{10}	1.036 (8.24 %)^{12}	0.948 (7.54 %)^{9}	0.885 (7.04 %)^{6}
Haliutu	2.242 (11.74 %)^{12}	2.138 (11.19 %)^{7}	2.228 (11.66 %)^{11}	2.199 (11.51 %)^{10}	2.155 (11.28 %)^{9}	2.127 (11.13 %)^{6}
Wulumuqi	0.938 (6.61 %)^{10}	0.936 (6.59 %)^{8}	0.942 (6.63 %)^{11}	0.957 (6.74 %)^{12}	0.937 (6.6 %)^{9}	0.935 (6.59 %)^{7}
Hetian	1.485 (9.57 %)^{11}	1.396 (9 %)^{7}	1.472 (9.49 %)^{10}	1.505 (9.7 %)^{12}	1.452 (9.36 %)^{9}	1.401 (9.03 %)^{8}
Jiuquan	0.768 (4.66 %)^{12}	0.698 (4.23 %)^{4}	0.715 (4.34 %)^{9}	0.734 (4.45 %)^{11}	0.711 (4.31 %)^{8}	0.696 (4.22 %)^{3}
Beijing	0.877 (6.59 %)^{12}	0.719 (5.4 %)^{4}	0.835 (6.27 %)^{11}	0.827 (6.21 %)^{10}	0.797 (5.99 %)^{9}	0.732 (5.5 %)^{8}
Xian	1.249 (10.83 %)^{12}	1.158 (10.04 %)^{6}	1.204 (10.44 %)^{9}	1.221 (10.59 %)^{10}	1.222 (10.6 %)^{11}	1.182 (10.25 %)^{8}
Geermu	0.917 (8.51 %)^{12}	0.623 (5.78 %)^{7}	0.792 (7.35 %)^{11}	0.788 (7.32 %)^{10}	0.785 (7.29 %)^{9}	0.642 (5.96 %)^{8}
Lasha	0.828 (4.04 %)^{12}	0.763 (3.73 %)^{5}	0.828 (4.04 %)^{11}	0.789 (3.85 %)^{6}	0.804 (3.93 %)^{10}	0.802 (3.92 %)^{9}
Chendu	1.19 (13.83 %)^{12}	1.104 (12.84 %)^{8}	1.139 (13.25 %)^{11}	1.139 (13.24 %)^{10}	1.121 (13.04 %)^{9}	1.07 (12.44 %)^{5}
Shanghai	0.914 (7.34 %)^{12}	0.718 (5.77 %)^{7}	0.845 (6.79 %)^{11}	0.838 (6.73 %)^{10}	0.828 (6.65 %)^{9}	0.772 (6.2 %)^{8}
Changsha	1.525 (14.4 %)^{9}	1.41 (13.32 %)^{2}	1.548 (14.62 %)^{11}	1.555 (14.69 %)^{12}	1.535 (14.5 %)^{10}	1.413 (13.34 %)^{3}
Kunming	1.458 (9.58 %)^{12}	1.301 (8.55 %)^{8}	1.395 (9.16 %)^{11}	1.365 (8.97 %)^{9}	1.385 (9.1 %)^{10}	1.247 (8.19 %)^{3}
Guangzhou	0.972 (8.3 %)^{11}	0.92 (7.86 %)^{8}	0.951 (8.12 %)^{9}	0.971 (8.29 %)^{10}	0.972 (8.3 %)^{12}	0.893 (7.62 %)^{3}
Sanya	1.491 (9.38 %)^{12}	1.428 (8.98 %)^{7}	1.476 (9.29 %)^{10}	1.477 (9.29 %)^{11}	1.475 (9.29 %)^{9}	1.384 (8.71 %)^{2}

Station	SVM7	SVM8	SVM9	SVM10	SVM11	SVM12
Haerbing	0.881 (7.01 %)^{4}	0.865 (6.88 %)^{1}	0.918 (7.31 %)^{8}	0.881 (7.02 %)^{5}	0.869 (6.92 %)^{2}	0.869 (6.92 %)^{3}
Haliutu	2.099 (10.99 %)^{5}	2.09 (10.94 %)^{3}	2.142 (11.21 %)^{8}	2.074 (10.86 %)^{1}	2.095 (10.97 %)^{4}	2.088 (10.93 %)^{2}
Wulumuqi	0.898 (6.32 %)^{4}	0.873 (6.15 %)^{2}	0.93 (6.55 %)^{5}	0.934 (6.58 %)^{6}	0.877 (6.18 %)^{3}	0.866 (6.1 %)^{1}
Hetian	1.349 (8.7 %)^{3}	1.353 (8.72 %)^{4}	1.391 (8.96 %)^{6}	1.382 (8.91 %)^{5}	1.328 (8.56 %)^{2}	1.326 (8.55 %)^{1}
Jiuquan	0.669 (4.06 %)^{2}	0.662 (4.02 %)^{1}	0.7 (4.25 %)^{5}	0.702 (4.26 %)^{6}	0.708 (4.3 %)^{7}	0.719 (4.36 %)^{10}
Beijing	0.726 (5.45 %)^{5}	0.729 (5.47 %)^{7}	0.707 (5.31 %)^{2}	0.727 (5.46 %)^{6}	0.703 (5.28 %)^{1}	0.71 (5.33 %)^{3}
Xian	1.168 (10.13 %)^{7}	1.147 (9.94 %)^{3}	1.138 (9.87 %)^{2}	1.15 (9.97 %)^{5}	1.149 (9.96 %)^{4}	1.128 (9.78 %)^{1}
Geermu	0.608 (5.64 %)^{3}	0.606 (5.63 %)^{1}	0.607 (5.64 %)^{2}	0.618 (5.73 %)^{6}	0.611 (5.67 %)^{4}	0.616 (5.72 %)^{5}
Lasha	0.801 (3.91 %)^{8}	0.801 (3.91 %)^{7}	0.738 (3.61 %)^{2}	0.75 (3.66 %)^{4}	0.747 (3.65 %)^{3}	0.737 (3.6 %)^{1}
Chendu	1.089 (12.66 %)^{7}	1.08 (12.55 %)^{6}	1.025 (11.92 %)^{3}	1.029 (11.97 %)^{4}	0.991 (11.52 %)^{1}	1.02 (11.86 %)^{2}
Shanghai	0.699 (5.62 %)^{4}	0.702 (5.64 %)^{5}	0.662 (5.32 %)^{1}	0.707 (5.68 %)^{6}	0.671 (5.39 %)^{3}	0.669 (5.38 %)^{2}
Changsha	1.46 (13.78 %)^{8}	1.442 (13.62 %)^{7}	1.403 (13.25 %)^{1}	1.414 (13.35 %)^{4}	1.423 (13.44 %)^{5}	1.425 (13.46 %)^{6}
Kunming	1.282 (8.42 %)^{7}	1.28 (8.41 %)^{6}	1.238 (8.14 %)^{2}	1.194 (7.85 %)^{1}	1.259 (8.27 %)^{4}	1.279 (8.41 %)^{5}
Guangzhou	0.913 (7.79 %)^{7}	0.879 (7.51 %)^{1}	0.912 (7.79 %)^{6}	0.885 (7.56 %)^{2}	0.91 (7.77 %)^{4}	0.911 (7.78 %)^{5}
Sanya	1.438 (9.05 %)^{8}	1.415 (8.9 %)^{6}	1.411 (8.88 %)^{5}	1.383 (8.7 %)^{1}	1.409 (8.87 %)^{4}	1.397 (8.79 %)^{3}

[a] RMSE values are given both in absolute values and in percentage, the superscript numbers represent for the RMSE based ranking of each model at different stations

kernel function $K(x_i,x_j)=\exp(-\gamma\|x_i, -x_j\|^2)$ is used, where d and γ are the kernel parameters. A total of 20 SVM models (Table 2) using different combinations of sunshine ratio, maximum and minimum air temperature, relative humidity, and atmospheric water vapor pressure as input attributes are explored and implemented by LIBSVM (Version 3.12) package, which is a library for SVM developed by Chang and Lin (2011). This package has helped users to easily apply SVM to applications and gained wide popularity in machine learning and many areas.

2.3 Empirical models

2.3.1 A-P model

Ångström (1924) proposed a simple linear relationship between the ratio of actual solar radiation to extra-terrestrial solar radiation and the ratio of actual sunshine duration to potential sunshine duration, which was further modified by Prescott (1940) with the following form:

$$\frac{H_s}{H_a} = a\frac{S}{S_o} + b \tag{7}$$

where H_s is actual solar radiation (MJ m^{-2}), H_a is extra-terrestrial solar radiation (MJ m^{-2}), S is the actual sunshine duration (h), S_o is potential sunshine duration (h), and a and b are empirical parameters which are calibrated from regression analysis between H_s/H_a and S/S_o. The extra-terrestrial solar radiation and potential sunshine duration are calculated using the equations detailed by Allen et al. (1998).

$$H_a = 37.6d(\omega\sin\varphi\sin\delta + \cos\varphi\cos\delta\sin\omega) \tag{8}$$

$$d = 1 + 0.033\cos\left(\frac{2\pi}{365}n\right) \tag{9}$$

$$\delta = 0.4093\sin\left(\frac{2\pi}{365}n - 1.39\right) \tag{10}$$

$$\omega = \arccos(-\tan\varphi\tan\delta) \tag{11}$$

$$S_o = 24\omega\Big/\pi \tag{12}$$

where d is the relative distance between the sun and the earth, ω is sunset hour angle (rad), φ is latitude (rad), δ is solar declination angle (rad), and n is the number of the day of year starting from the first of January.

2.3.2 H-S model

Hargreaves et al. (1985) suggested a simple equation for estimation of solar radiation using air temperature range (the difference between maximum and minimum air temperature) to solve the problem of availability of sunshine duration data.

$$\frac{H_s}{H_a} = a\sqrt{T_{max} - T_{min}} + b \tag{12}$$

Table 4 Bayesian information criterion of sunshine duration-based SVM models[a]

Station	SVM1	SVM2	SVM3	SVM4	SVM5	SVM6	SVM7	SVM8	SVM9	SVM10	SVM11	SVM12
Haerbing	4.50^{9}	-1.37^{1}	8.22^{11}	13.78^{12}	-0.33^{3}	-0.90^{2}	3.12^{6}	1.70^{4}	4.11^{8}	3.99^{7}	2.32^{5}	7.13^{10}
Haliutu	47.59^{1}	48.99^{2}	51.22^{5}	50.53^{4}	52.71^{7}	51.99^{6}	49.44^{3}	55.75^{10}	53.09^{8}	55.34^{9}	55.88^{11}	59.71^{12}
Wulumuqi	-2.84^{1}	1.62^{3}	2.36^{5}	4.27^{8}	6.30^{10}	1.38^{2}	1.78^{4}	2.82^{6}	5.59^{9}	11.02^{12}	3.42^{7}	6.63^{11}
Hetian	52.27^{3}	49.62^{1}	55.98^{9}	58.63^{12}	54.84^{7}	50.32^{2}	54.32^{6}	55.39^{8}	53.93^{5}	58.00^{11}	53.22^{4}	57.83^{1}
Jiuquan	-11.73^{5}	-13.41^{1}	-11.96^{3}	-10.35^{06}	-9.48^{7}	-11.80^{4}	-12.28^{2}	-8.36^{9}	-9.08^{8}	-4.84^{10}	-4.32^{11}	0.69^{12}
Beijing	-10.97^{12}	-29.96^{1}	-12.11^{11}	-13.27^{10}	-23.03^{5}	-24.10^{3}	-17.65^{8}	-18.80^{7}	-27.22^{2}	-19.17^{6}	-23.13^{4}	-17.12^{9}
Xian	16.94^{2}	16.61^{1}	18.87^{3}	19.71^{5}	21.86^{8}	21.18^{7}	19.76^{6}	24.18^{9}	19.66^{4}	24.34^{11}	24.28^{10}	27.30^{12}
Geermu	-5.66^{12}	-47.25^{1}	-18.46^{11}	-19.01^{10}	-38.75^{6}	-45.38^{3}	-19.48^{9}	-40.96^{4}	-45.49^{2}	-38.66^{7}	-40.02^{5}	-34.21^{8}
Lasha	-17.84^{4}	-22.93^{1}	-13.06^{8}	-18.87^{3}	-12.16^{11}	-12.2^{1}	-16.57^{5}	-7.54^{12}	-22.05^{2}	-15.40^{7}	-15.92^{6}	-12.69^{9}
Chendu	23.46^{10}	20.06^{3}	23.45^{9}	23.42^{8}	21.34^{4}	23.25^{7}	21.71^{5}	27.01^{12}	16.71^{1}	21.85^{6}	17.75^{2}	25.53^{11}
Shanghai	-1.30^{12}	-11.68^{2}	-1.91^{11}	-2.45^{10}	-3.23^{8}	-9.18^{3}	-3.15^{9}	-4.87^{5}	-12.44^{1}	-4.42^{6}	-7.51^{4}	-3.62^{7}
Changsha	34.65^{2}	33.30^{1}	39.99^{6}	40.34^{8}	37.72^{4}	40.05^{7}	39.42^{5}	43.48^{11}	37.20^{3}	42.04^{9}	42.52^{10}	46.90^{12}
Kunming	49.30^{11}	40.60^{4}	48.79^{10}	46.21^{7}	40.34^{3}	43.63^{5}	47.93^{8}	48.16^{9}	39.53^{1}	40.04^{2}	46.21^{6}	52.91^{12}
Guangzhou	1.38^{3}	-0.43^{1}	3.57^{6}	5.98^{9}	0.73^{2}	3.39^{5}	6.19^{10}	3.71^{7}	3.28^{4}	4.51^{8}	7.77^{11}	12.70^{12}
Sanya	28.06^{1}	29.55^{2}	31.56^{4}	31.58^{5}	31.79^{6}	34.07^{8}	31.52^{3}	37.19^{11}	32.93^{7}	35.81^{9}	36.93^{1}	40.53^{12}
Overall ranking	4	1	8	9	5	6	3	11	2	10	7	12

[a] The superscript numbers represent for the Bayesian information criterion-based ranking of each model at different station

where $T_{\max}$ and $T_{\min}$ are the maximum and minimum air temperature, respectively, and a and b are empirical parameters.

2.3.3 Chen model

Chen et al. (2004) modified A-P model by introducing air temperature range and claimed better results at 32 stations all over China:

$$\frac{H_s}{H_a} = a\frac{S}{S_o} + b\ln(T_{\max} - T_{\min}) + c \tag{13}$$

where a, b and c are empirical parameters.

2.3.4 Local models

Except these well-known models, two equations were created using the training data sets through multiple regression analysis (hereafter local1 [Eq. 14] and local2 [Eq. 15]).

$$\frac{H}{H_a} = a\sqrt{T_{\max} - T_{\min}} + b\mathrm{RH} + c\mathrm{VP} + d \tag{14}$$

$$\frac{H}{Ha} = a\frac{S}{S_o} + b\sqrt{(T_{\max} - T_{\min})} + c\mathrm{RH} + d\mathrm{VP} + e \tag{15}$$

where RH and VP are monthly mean relative humidity and atmospheric water vapor pressure, respectively, and a, b, c, d and e are empirical parameters.

2.4 Performance criteria

Root mean square error (RMSE) and determination coefficient (R^2) are used to evaluate model performance. RMSE provides information on the short term performance of the correlations by allowing a term by term comparison of the actual deviation between the estimated and measured values. Lower values of RMSE indicate a better performance. In order to give a general judgement of the model performance at different station, RMSE values are also expressed in percentage (%) which are normalized by the respective measured means. The metric R^2 varying between 0 and 1 is adopted to measure the fit of model, higher the value, better the fit. RMSE is calculated by the following equations.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}\left(y_i - \widehat{y}_i\right)^2}{n}} \tag{16}$$

where n, y, and $\widehat{y}$ represent the total number of evaluating data, the observation, and the estimation, respectively.

When developing models, it is possible to increase the likelihood by adding more parameters, but doing so may

Table 5 Root mean square errors (RMSE) of air temperature-based SVM models[a]

Station	SVM13	SVM14	SVM15	SVM16	SVM17	SVM18	SVM19	SVM20
Haerbing	1.156 (9.21 %)^{7}	1.18 (9.39 %)^{8}	1.097 (8.73 %)^{5}	1.051 (8.36 %)^{2}	1.093 (8.7 %)^{4}	1.129 (8.99 %)^{6}	1.038 (8.26 %)^{1}	1.053 (8.38 %)^{3}
Haliutu	2.599 (13.61 %)^{8}	2.575 (13.49 %)^{7}	2.343 (12.27 %)^{2}	2.348 (12.3 %)^{3}	2.461 (12.88 %)^{6}	2.431 (12.73 %)^{5}	2.36 (12.35 %)^{4}	2.329 (12.19 %)^{1}
Wulumuqi	1.261 (8.88 %)^{8}	1.213 (8.54 %)^{6}	1.172 (8.26 %)^{4}	1.139 (8.03 %)^{3}	1.217 (8.57 %)^{7}	1.196 (8.43 %)^{5}	1.114 (7.85 %)^{1}	1.115 (7.86 %)^{2}
Hetian	1.719 (11.08 %)^{8}	1.695 (10.93 %)^{7}	1.58 (10.19 %)^{4}	1.551 (10 %)^{2}	1.688 (10.88 %)^{6}	1.663 (10.71 %)^{5}	1.558 (10.04 %)^{3}	1.541 (9.93 %)^{1}
Jiuquan	1.115 (6.77 %)^{7}	1.117 (6.77 %)^{8}	0.842 (5.1 %)^{3}	0.844 (5.12 %)^{4}	1.099 (6.67 %)^{5}	1.113 (6.75 %)^{6}	0.83 (5.03 %)^{2}	0.827 (5.01 %)^{1}
Beijing	1.395 (10.47 %)^{8}	1.356 (10.18 %)^{6}	1.201 (9.02 %)^{3}	1.177 (8.84 %)^{1}	1.362 (10.23 %)^{7}	1.329 (9.98 %)^{5}	1.226 (9.2 %)^{4}	1.181 (8.87 %)^{2}
Xian	1.324 (11.49 %)^{8}	1.311 (11.37 %)^{7}	1.248 (10.83 %)^{4}	1.187 (10.3 %)^{3}	1.31 (11.36 %)^{6}	1.309 (11.36 %)^{5}	1.135 (9.84 %)^{1}	1.136 (9.85 %)^{2}
Geermu	1.132 (10.52 %)^{8}	1.104 (10.26 %)^{7}	1.026 (9.53 %)^{5}	0.978 (9.08 %)^{3}	1.028 (9.54 %)^{6}	0.997 (9.25 %)^{4}	0.971 (9.01 %)^{2}	0.969 (8.99 %)^{1}
Lasha	1.395 (6.81 %)^{7}	1.415 (6.91 %)^{8}	1.257 (6.14 %)^{4}	1.211 (5.91 %)^{3}	1.329 (6.49 %)^{6}	1.322 (6.45 %)^{5}	1.204 (5.88 %)^{2}	1.203 (5.87 %)^{1}
Chendu	1.547 (18 %)^{8}	1.535 (17.84 %)^{6}	1.47 (17.1 %)^{4}	1.469 (17.08 %)^{3}	1.529 (17.78 %)^{5}	1.538 (17.89 %)^{7}	1.404 (16.31 %)^{1}	1.425 (16.57 %)^{2}
Shanghai	1.692 (13.6 %)^{8}	1.686 (13.55 %)^{7}	1.435 (11.52 %)^{4}	1.394 (11.19 %)^{2}	1.611 (12.94 %)^{5}	1.618 (12.99 %)^{6}	1.416 (11.37 %)^{3}	1.371 (11 %)^{1}
Changsha	2.212 (20.89 %)^{7}	2.255 (21.29 %)^{8}	1.859 (17.56 %)^{4}	1.824 (17.23 %)^{3}	2.08 (19.65 %)^{6}	2.06 (19.46 %)^{5}	1.747 (16.49 %)^{1}	1.799 (16.98 %)^{2}
Kunming	1.883 (12.38 %)^{8}	1.853 (12.18 %)^{7}	1.706 (11.2 %)^{4}	1.673 (10.99 %)^{3}	1.746 (11.48 %)^{6}	1.715 (11.26 %)^{5}	1.623 (10.66 %)^{1}	1.639 (10.76 %)^{2}
Guangzhou	1.667 (14.23 %)^{8}	1.65 (14.1 %)^{7}	1.567 (13.39 %)^{4}	1.529 (13.06 %)^{3}	1.579 (13.49 %)^{6}	1.567 (13.39 %)^{5}	1.482 (12.66 %)^{1}	1.508 (12.88 %)^{2}
Sanya	1.984 (12.49 %)^{8}	1.959 (12.33 %)^{7}	1.791 (11.27 %)^{4}	1.781 (11.2 %)^{3}	1.824 (11.48 %)^{6}	1.795 (11.29 %)^{5}	1.747 (10.99 %)^{1}	1.749 (11 %)^{2}

[a] RMSE values are given both in absolute values and in percentage, the superscript numbers represent for the RMSE-based ranking of each model at different stations

result in parsimoniousness. Bayesian information criterion (BIC) resolves this problem by introducing a penalty term for the number of parameters. It has been widely used to select the optimal model identified by the minimum value.

$$\mathrm{BIC} = n\ln\left(\frac{\mathrm{RSS}}{n}\right) + k\ln(n) \tag{17}$$

where k and RSS are the number of parameters and the residual sum of squares from the model, respectively.

3 Results

3.1 Performances of SVM models

3.1.1 Sunshine duration data are available

Performances of the developed SVM models are presented in Tables 3, 4, 5 and 6. When sunshine duration data are available, SVM models (SVM1–12) give good performances with RMSE <2.3 MJ m^{-2} (15 %) and the average RMSE of 1.097 MJ m^{-2} (8.1 %). On average, SVM12 yields the lowest RMSE of 1.05 MJ m^{-2} (7.8 %), while SVM6–11 have a very similar RMSE values to SVM12. Obviously, models perform differently at different stations. Each of them is ranked based on RMSE for each station and presented in Table 3. SVM8 performs best at Geermu, Haerbing, Guangzhou and Jiuquan, SVM9 at Changsha and Shanghai, SVM10 at Hailiutu, Kunming and Sanya, SVM11 at Beijing and Chengdu, and SVM12 at Hetian, Lasha, Wulumuqi and Xian.

SVM8–12 seem the best at the corresponding stations, while they may need more input parameters, for example, SVM12 needs five, SVM10–11 have four, and only one for SVM1. When the parsimoniousness of the models is considered by taking into account the number of input parameters (Table 4), SVM2 using sunshine ratio and air temperature range performs best at nine stations, SVM9 using sunshine ratio, maximum and minimum air temperature performs best at three stations, and SVM1 using sunshine ratio only at the remainder (Table 4). According to the overall ranking, SVM2 is superior to other models.

SVM2 and SVM9 have an average 10 % and 12 % lower RMSE than SVM1, respectively; suggesting that additional input of air temperature can significantly improve the accuracy of SVM model which uses sunshine ratio only. This is also indicated by the lower RMSE (average 12 %) of SVM6–8, SVM10–12 than SVM1. SVM3–4 give similar performances to SVM1, suggesting that atmospheric water vapor pressure and relative humidity do not adequately account for the improvement in estimation accuracy. The similar average RMSE of SVM6–8 to those of SVM2 further confirms this result. Therefore, if all of those meteorological variables are available, it is unnecessary to take into account atmospheric water vapor pressure and relative humidity. SVM2 is proposed and can provide a good method for estimation of monthly mean daily solar radiation.

3.1.2 Sunshine duration data are not available

When sunshine duration data are not available, SVM models (SVM13–20) return reasonable results (Table 5) with RMSE <2.6 MJ m^{-2} (22 %) and the average RMSE of 1.485 MJ m^{-2} (11.1 %). SVM20, which performs best at Geermu, Hailiutu, Hetian, Jiuquan, Lasha and Shanghai, yields lowest average RMSE of 1.389 MJ m^{-2} (10.4 %)

Table 6 Bayesian information criterion of air temperature based SVM models[a]

Station	SVM13	SVM14	SVM15	SVM16	SVM17	SVM18	SVM19	SVM20
Haerbing	22.23[5]	29.40[8]	20.64[4]	20.29[3]	20.24[2]	28.96[7]	18.85[1]	25.33[6]
Haliutu	55.56[2]	59.05[6]	53.96[1]	58.06[4]	56.60[3]	59.92[7]	58.32[5]	61.61[8]
Wulumuqi	32.59[5]	32.74[6]	28.62[2]	29.99[3]	33.14[7]	35.84[8]	27.34[1]	32.22[4]
Hetian	69.80[4]	72.92[7]	64.47[1]	67.04[2]	72.38[6]	75.37[8]	67.54[3]	71.01[5]
Jiuquan	10.63[5]	14.81[7]	−2.14[1]	2.08[3]	13.85[6]	18.70[8]	1.07[2]	4.96[4]
Beijing	44.70[5]	46.11[6]	31.57[1]	33.93[2]	46.68[7]	48.49[8]	38.79[3]	39.11[4]
Xian	20.34[2]	23.82[7]	20.98[3]	22.14[4]	23.76[6]	27.81[8]	19.53[1]	23.63[5]
Geermu	19.67[7]	21.49[8]	12.65[3]	11.71[2]	12.88[4]	13.98[5]	10.80[1]	15.36[6]
Lasha	44.70[6]	51.22[8]	37.04[2]	37.31[3]	43.66[5]	47.85[7]	36.59[1]	41.32[4]
Chendu	51.80[3]	55.64[6]	50.98[2]	55.59[5]	55.23[4]	60.56[8]	50.66[1]	56.98[7]
Shanghai	35.65[5]	39.53[7]	29.83[1]	32.20[2]	36.79[6]	41.15[8]	33.14[3]	35.29[4]
Changsha	61.42[6]	67.09[8]	53.20[2]	56.11[3]	61.29[5]	64.87[7]	52.99[1]	59.374
Kunming	79.47[7]	82.33[8]	72.56[2]	75.06[3]	75.32[4]	77.94[6]	71.44[1]	77.35[5]
Guangzhou	66.10[5]	69.69[8]	63.47[2]	65.34[4]	64.39[3]	68.28[6]	61.58[1]	68.45[7]
Sanya	45.19[3]	48.54[7]	43.15[1]	46.91[5]	44.23[2]	47.38[6]	45.74[4]	49.91[8]
Overall ranking	4	7	2	3	5	8	1	6

[a]The superscript numbers represent for Bayesian information criterion based ranking of each model at different stations

followed by SVM19 which performs best at Changsha, Chengdu, Guangzhou, Haerbing, Kunming, Sanya, Wulumuqi and Xian. When the parsimoniousness of SVM13–20 is considered (Table 6), SVM models (SVM15 and 19) using air temperature in combination with atmospheric water vapor pressure perform better than other models. According to overall ranking (Table 6), SVM19 is the optimal model followed by SVM15.

SVM15 and 19 are superior to SVM13 and 17 with an average of 10 % higher accuracy, suggesting that an additional input of atmospheric water vapor pressure can significantly improve the accuracy of the SVM models which uses air temperature only. This is also indicated by the lower RMSE (average 10 %) of SVM16 and 20 than SVM14 and 18. While SVM 14 and 18 give similar performances to SVM 13 and 17, suggesting that relative humidity does not contribute to the improvement in estimation accuracy. The similar RMSE of SVM 20 to those of 19 further confirms this result, and therefore additional inclusion of relative humidity is unnecessary.

Table 7 Calibrated parameters of the study empirical models

Station	A-P[a]			H-S[b]			Chen[c]			
	a	*b*	R^2	*a*	*b*	R^2	*a*	*b*	*c*	R^2
Haerbing	0.456	0.240	0.502	0.122	0.095	0.378	0.384	0.102	0.034	0.534
Haliutu	0.510	0.249	0.580	0.081	0.322	0.391	0.505	0.006	0.236	0.581
Wulumuqi	0.511	0.209	0.788	0.213	−0.181	0.504	0.450	0.069	0.083	0.799
Hetian	0.430	0.296	0.675	0.111	0.165	0.381	0.401	0.058	0.169	0.692
Jiuquan	0.402	0.316	0.561	0.100	0.227	0.368	0.377	0.078	0.129	0.584
Beijing	0.601	0.141	0.741	0.184	−0.088	0.420	0.544	0.055	0.046	0.749
Xian	0.416	0.235	0.653	0.179	−0.177	0.533	0.298	0.130	−0.019	0.718
Geermu	0.572	0.255	0.707	0.179	−0.006	0.405	0.531	0.128	−0.052	0.732
Lasha	0.499	0.297	0.442	0.190	−0.076	0.412	0.504	−0.005	0.307	0.442
Chendu	0.563	0.161	0.788	0.210	−0.262	0.650	0.427	0.093	0.012	0.814
Shanghai	0.550	0.172	0.789	0.140	0.047	0.385	0.531	0.029	0.126	0.793
Chasha	0.612	0.127	0.869	0.275	−0.411	0.516	0.588	0.025	0.086	0.870
Kunming	0.522	0.190	0.806	0.200	−0.185	0.715	0.488	0.024	0.152	0.807
Guangzhou	0.554	0.137	0.854	0.345	−0.588	0.701	0.419	0.156	−0.122	0.881
Sanya	0.352	0.300	0.545	0.131	0.157	0.384	0.327	0.117	0.098	0.636

Station	Local1[d]					Local2[e]					
	a	*b*	*c*	*d*	R^2	*a*	*b*	*c*	*d*	*e*	R^2
Haerbing	0.113	−0.022	0.031	−2.974	0.449	0.369	0.056	−0.015	0.029	−2.748	0.610
Haliutu	0.099	−0.056	0.068	−5.633	0.499	0.374	0.022	−0.062	0.042	−3.382	0.692
Wulumuqi	0.062	−0.353	0.017	−1.034	0.666	0.382	0.020	−0.123	0.008	−0.467	0.818
Hetian	0.214	0.086	0.097	−8.640	0.595	0.293	0.097	0.003	0.048	−4.111	0.745
Jiuquan	0.147	−0.034	0.060	−5.046	0.437	0.314	0.070	0.000	0.018	−1.385	0.598
Beijing	0.187	−0.001	0.035	−3.605	0.623	0.463	0.071	0.055	0.018	−1.865	0.779
Xian	0.144	−0.142	−0.009	0.932	0.552	0.295	0.058	−0.124	−0.004	0.560	0.729
Geermu	0.175	−0.093	0.113	−8.129	0.492	0.469	0.077	−0.041	0.028	−1.969	0.744
Lasha	0.197	−0.011	0.048	−3.243	0.426	0.513	0.022	0.070	−0.023	1.706	0.445
Chendu	0.167	−0.006	−0.037	3.356	0.752	0.363	0.052	−0.151	−0.021	2.219	0.840
Shanghai	−0.005	−1.020	−0.046	5.846	0.489	0.466	−0.008	−0.330	−0.021	2.644	0.830
Chasha	0.171	−0.448	−0.051	5.353	0.693	0.554	0.039	0.112	−0.013	1.259	0.883
Kunming	0.134	−0.333	0.041	−3.051	0.738	0.477	−0.031	−0.250	0.031	−2.030	0.819
Guangzhou	0.271	−0.459	−0.050	5.071	0.773	0.400	0.086	−0.185	−0.008	0.931	0.887
Sanya	0.266	0.001	−0.064	6.251	0.424	0.319	0.139	0.194	−0.002	−0.026	0.659

[a] Ångström (1924), Prescott (1940)

[b] Hargreaves et al. (1985)

[c] Chen et al. (2004)

[d] Local regressive model (Eq. 14)

[e] Local regressive model (Eq. 15)

3.2 Performance of the empirical models

The locally calibrated parameters and performances of the empirical models are presented in Tables 7 and 8, respectively. All the empirical models perform well with RMSE varying between 0.886 (4.6 %) and 2.675 MJ m^{-2} (21.5 %). Models including sunshine ratio (A-P, Chen and Local2) perform much better than models including air temperature (H-S and Local1) with 7–40 % (average 23 %) higher accuracy. On average, Local2 model has the lowest RMSE of 1.297 MJ m^{-2} (9.7 %). While Chen (RMSE=1.309 MJ m^{-2} (9.8 %)) and A-P (RMSE=1.343 MJ m^{-2} (10 %)) models give similar performances to Local2. This indicates that air temperature range, atmospheric water vapor pressure and relative humidity, as introduced in an additive form, do not adequately account for the improvement in accuracy of the A-P model. Our results are different from those reported by Chen et al. (2004), who introduced air temperature range to A-P model and claimed a better performance, but are consistent with those of Wu et al. (2007), who reported similar performances between Chen and the original A-P model. Local1 model is superior to the widely used H-S model with 4–16 % (average 8 %) higher accuracy, indicating that additional inclusion of atmospheric water vapor pressure and relative humidity can significantly improve the accuracy of the temperature-based models.

3.3 Comparison of SVM between empirical models

To demonstrate the potential of SVM, further comparisons of the results of SVM with those of empirical models are made. SVM1, SVM2 and SVM8 show a 4–33 % (average 16 %) lower RMSE than the corresponding A-P, Chen and Local2 models. This illustrates that SVM significantly outperforms the empirical model. The lower RMSE (average 11 %) of SVM13 and 16 than the corresponding H-S and Local1 models confirms this result. When sunshine duration data are available, the optimal model (SVM2) shows a 13–35 % (average 21 %), 20–59 % (average 39 %), 11–33 % (average 19 %), 17–55 % (average 34 %), and 10–33 % (average 18 %) lower RMSE (average 26 %) than A-P, H-S, Chen, Local1, and Local2 models, respectively. When sunshine duration data are not available, the optimal model (SVM19) shows a 12–35 % (average 21 %) and 8–24 % (average 15 %) lower RMSE (average 18 %) than H-S and Local1 models, respectively. These results further prove the superiority of SVM over the empirical models. The remarkable improvement indicates that SVM method would be a promising alternative over the traditional approaches for estimation of solar radiation at any locations.

Table 8 Root mean square errors (RMSE) of the study empirical models

Station	A-P[a]	H-S[b]	Chen[c]	Local1[d]	Local2[e]
Haerbing	1.141 (9.09 %)	1.387 (11.04 %)	1.105 (8.79 %)	1.281 (10.2 %)	1.136 (9.05 %)
Haliutu	2.51 (13.14 %)	2.675 (14 %)	2.389 (12.51 %)	2.567 (13.44 %)	2.369 (12.4 %)
Wulumuqi	1.098 (7.74 %)	1.507 (10.62 %)	1.059 (7.46 %)	1.381 (9.73 %)	1.049 (7.39 %)
Hetian	1.686 (10.87 %)	1.896 (12.22 %)	1.596 (10.29 %)	1.784 (11.5 %)	1.568 (10.1 %)
Jiuquan	0.936 (5.68 %)	1.281 (7.77 %)	0.935 (5.67 %)	1.09 (6.61 %)	0.886 (5.37 %)
Beijing	1.019 (7.66 %)	1.581 (11.87 %)	1.019 (7.65 %)	1.388 (10.42 %)	0.995 (7.47 %)
Xian	1.333 (11.56 %)	1.513 (13.12 %)	1.333 (11.56 %)	1.456 (12.63 %)	1.319 (11.44 %)
Geermu	0.953 (8.85 %)	1.427 (13.25 %)	0.935 (8.68 %)	1.206 (11.19 %)	0.925 (8.59 %)
Lasha	0.956 (4.67 %)	1.454 (7.1 %)	0.933 (4.56 %)	1.367 (6.68 %)	0.937 (4.58 %)
Chendu	1.379 (16.04 %)	1.676 (19.49 %)	1.338 (15.55 %)	1.607 (18.68 %)	1.345 (15.63 %)
Shanghai	1.026 (8.24 %)	1.77 (14.22 %)	1.005 (8.07 %)	1.591 (12.78 %)	1.012 (8.13 %)
Changsha	1.747 (16.49 %)	2.272 (21.46 %)	1.703 (16.08 %)	2.075 (19.6 %)	1.674 (15.81 %)
Kunming	1.605 (10.54 %)	1.949 (12.8 %)	1.572 (10.33 %)	1.857 (12.2 %)	1.569 (10.31 %)
Guangzhou	1.117 (9.54 %)	1.767 (15.09 %)	1.121 (9.57 %)	1.697 (14.49 %)	1.08 (9.22 %)
Sanya	1.642 (10.33 %)	2.062 (12.98 %)	1.596 (10.04 %)	1.912 (12.03 %)	1.591 (10.02 %)

RMSE values are given both in absolute values and in percentage

[a] Ångström (1924), Prescott (1940)

[b] Hargreaves et al. (1985)

[c] Chen et al. (2004)

[d] Local regressive model (Eq. 14)

[e] Local regressive model (Eq. 15)

4 Conclusions

The feasibility and potential of SVM in estimation of solar radiation are investigated in this work. Twenty SVM models using different combinations of sunshine ratio, maximum and minimum air temperature, relative humidity, and atmospheric water vapor pressure as input attributes are explored. All the SVM models give good performances and significantly outperform the empirical models. When sunshine duration data are available, input of air temperature significantly improves the accuracy of the SVM models which uses sunshine ratio only. While atmospheric water vapor pressure and relative humidity do not adequately account for improvement in estimation accuracy. The optimal model (SVM2) using sunshine ratio and air temperature range is proposed. When sunshine duration data are not available, input of atmospheric water vapor pressure significantly improves the accuracy of the SVM models which uses air temperature only. While relative humidity does not contribute to the improvement in estimation accuracy. The optimal model (SVM19) using maximum temperature, minimum temperature and atmospheric water vapor pressure is proposed in this case. The remarkable improvement indicates that the SVM method would be a promising alternative over the traditional approaches for estimation of solar radiation at any locations.

Acknowledgments The work was supported by National Key Technology Research and Development Program (2012BAC21B01), the Geological Survey program of China Geological Survey (GZH201200503) and Special foundation for scientific research on public Interest (1212010611402, 201111023). We thank the National Meteorological Information Center, China Meteorological Administration for providing the long-term data records. Many thanks go to the anonymous reviewers for the comments on the manuscript.

References

Abdalla YAG (1994) New correlation of global solar radiation with meteorological parameters for Bahrain. Int J Sol Energy 16:111–120

Allen RG, Pereira LS, Raesm D, Smith M (1998) Crop evapotranspiration—guidelines for computing crop water requirements. FAO Irrigation and Drainage Paper 56. Food and Agriculture Organization of the United Nations, Rome

Ångström A (1924) Solar and terrestrial radiation. Q J R Meteorol Soc 50:121–126

Ball RA, Purcell LC, Carey SK (2004) Evaluation of solar radiation prediction models in North America. Agron J 96:391–397

Bechini L, Ducco G, Donatelli M, Stein A (2000) Modelling, interpolation and stochastic simulation in space and time of global solar radiation. Agric Ecosyst Environ 81:29–42

Chang CC, Lin CJ (2011) LIBSVM: a library for support vector machines. ACM Trans Intell Syst Technol 2(3):1–27

Chen JL, Li GS (2012a) Assessing effect of time scale on the solar radiation–sunshine duration relationship. IDOJARAS 116(2):123–143

Chen JL, Li GS (2012b) Estimation of monthly average daily solar radiation from measured meteorological data in Yangtze River Basin in China. Int J Climatol 2013(33):487–498

Chen RS, Ersi K, Yang JP, Lu SH, Zhao WZ (2004) Validation of five global radiation models with measured daily data in China. Energy Convers Manag 45:1759–1769

Chen RS, Ersi K, Ji XB, Yang JP, Zhao WZ (2006) Trends of the global radiation and sunshine hours in 1961–1998 and their relationships in China. Energy Convers Manag 47:2859–2866

Chen JL, Liu HB, Wu W, Xie DT (2011) Estimation of monthly solar radiation from measured temperatures using support vector machines—a case study. Renew Energy 36(1):413–420

Choisnel E, de Villele O, Lacroze F (1992) Une aproche uniformisée du calcul de Ïévaporation potentielle por lénsemble de pays de la communauté européene. Publication EUR 14223 FR of the Office for Official Publications of the EU, Luxemburg

De JR, Stewart DW (1993) Estimating global solar radiation from common meteorological observations in western Canada. Can J Plant Sci 73:509–518

Dibike BY, Velickov S, Solomatine D, Abbott MB (2001) Model induction with support vector machines: introduction and applications. J Comput Civil Eng 15(3):208–216

Ducco G, Bechini L, Donatelli M, Marleto V (1998) Estimation and spatial interpolation of solar radiation in the Po valley, Italy. In: Proceedings of the Fifth ESA Conference, Nitra, pp 139–140

Ertekin C, Yaldiz O (1999) Estimation of monthly average daily global radiation on horizontal surface for Antalya, Turkey. Renew Energy 17:95–102

Garcia AGY, Hoogenboom G (2005) Evaluation of an improved daily solar radiation generator for the southeastern USA. Clim Res 29:91–102

Garg HP, Garg ST (1982) Prediction of global solar radiation from bright sunshine hours and other meteorological parameters. In: Solar-India, Proceedings of the National Solar Energy convention. Allied Publishers, New Delhi, pp 1004–1007

Gopinathan KK (1988) A general formula for computing the coefficients of the correlations connecting global solar radiation to sunshine duration. Sol Energy 41:499–502

Gueymard C, Jidra P, Eatrada CV (1995) A critical look at recent interpretations of the Ångström approach and its future in global solar irradiation prediction. Sol Energy 54:357–363

Hansen JW (1999) Stochastic daily solar irradiance for biological modeling applications. Agric For Meteorol 94:53–63

Hargreaves GL, Hargreaves GH, Riley JP (1985) Irrigation water requirement for Senegal River Basin. J Irrig Drain Eng ASAE 111:265–275

Hunt LA, Kuchar L, Swanton CJ (1998) Estimation of solar radiation for use in crop modelling. Agric For Meteorol 91:293–300

Iziomon MG, Mayer H (2002) Assessment of some global solar radiation parameterizations. J Atmos Sol Terr Phys 64:1631–1643

Janjai S, Pankaew P, Laksanaboonsong J (2009) A model for calculating hourly global solar radiation from satellite data in the tropics. Appl Energy 86:1450–1457

Kecman V (2005) Support vector machines—an introduction. In: Wang L (ed) Support vector machines: theory and applications. Springer-Verlag, New York, pp 1–48

Kuye A, Jagtap SS (1994) Correlation of solar radiation with climatological data for Port Harcourt, Nigeria. Int J Climatol 14:815–825

Lee SW, Verri A (2003) Support vector machines for computer vision and pattern recognition. Int J Pattern Recognit Artif Intell 3:331–332

Liong SY, Sivapragasam C (2002) Flood stage forecasting with support vector machines. J Am Water Resour Assoc 38:173–186

Liu DL, Scott BJ (2001) Estimation of solar radiation in Australia from rainfall and temperature observations. Agric For Meteorol 106:41–59

Liu XY, Mei XR, Li YZ, Wang QS, Jensen JR, Zhang XQ (2009a) Evaluation of temperature-based global solar radiation models in China. Agric Forest Meteorol 149:1433–1446

Liu XY, Mei XR, Li YZ, Zhang YQ, Wang QS (2009b) Calibration of the Ångström–Prescott coefficients (*a*, *b*) under different time scales and their impacts in estimating global solar radiation in the Yellow River basin. Agric For Meteorol 149:697–710

Lu WZ, Wang W (2005) Potential assessment of the support vector machine method in forecasting ambient air pollutant trends. Chemosphere 59:693–701

Manual UK, Victor US, Rao KV, Srivastava NN, Vittal KPR (2003) Estimation of solar radiation from temperature and rainfall Observations. J Agrometeorol 5:15–26

Mossad EM (2005) Sunshine and global solar radiation estimation at different sites in Egypt. J Atmos Sol Terr Phys 67:1331–1342

Ododo JC, Sulaiman AT, Aidan J, Yguda MM, Ogbu FA (1995) The importance of maximum air temperature in the parameterization of solar radiation in Nigeria. Renew Energy 6:751–763

Ojosu JO, Komolafe LK (1987) Models for estimating solar radiation availability in South Western Nigeria. Niger J Sol Energy 6:69–77

Pinker RT, Frouin R, Li Z (1995) A review of satellite methods to derive shortwave irradiance. Remote Sens Environ 5:108–124

Podestá GP, Núñez L, Villanueva CA, Skansi MA (2004) Estimating daily solar radiation in the Argentine Pampas. Agric For Meteorol 123:41–53

Polo J, Zarzalejo LF, Cony M, Navarro AA, Marchante R, Martin L, Romero M (2011) Solar radiation estimations over India using Meteosat satellite images. Solar Energy 85:2395–2406

Prescott JA (1940) Evaporation from a water surface in relation to solar radiation. Trans R Soc S Aust 64:114–118

Richardson CW (1981) Stochastic simulation of daily precipitation, temperature, and solar radiation. Water Resour Res 17:182–190

Rivington M, Bellocchi G, Matthews KB, Buchan K (2005) Evaluation of three model estimations of solar radiation at 24 UK stations. Agric For Meteorol 132:228–243

Schillings C, Mannstein H, Meyer R (2004) Operational method for deriving high resolution direct normal irradiance from satellite data. Sol Energy 76:475–484

Şenkal O (2010) Modeling of solar radiation using remote sensing and artificial neural network in Turkey. Energy 35:4795–4801

Soltani A, Meinke H, De VP (2003) Assessing linear interpolation to generate daily radiation and temperature data for use in crop simulations. Eur J Agron 21:133–158

Stanislaw O, Konrad G (2007) Forecasting of the daily meteorological pollution using wavelets and support vector machine. Eng Appl Artif Intell 20:745–755

Thornton PE, Running SW (1999) An improved algorithm for estimating daily solar radiation from measurements of temperature, humidity, and precipitation. Agric For Meteorol 93:211–228

Tirusew A, Mariush K, Mac M, Abedalrazq K (2006) Multi-time scale stream flow predictions: the support vector machines approach. J Hydrol 318:7–16

Trnka M, Zalud Z, Eitzinger J, Dubrovský M (2005) Global solar radiation in Central European lowlands estimated by various empirical formulae. Agric For Meteorol 131:54–76

Vapnik VN (1995) The nature of statistical learning theory. Springer Verlag, New York

Vapnik VN (1998) Statistical learning theory. Wiley, New York

Vapnik VN, Golowich SE, Smola AJ (1996) Support vector method for function approximation, regression estimation and signal processing. In: Mozer M, Jordan M, Petsche T (eds) Advances in neural information processing systems. MIT Press, Massachusetts, pp 281–287

Vignola F, Harlan P, Perez R, Kmiecik M (2007) Analysis of satellite derived beam and global solar radiation data. Sol Energy 81:768–772

Wallis TWR, Griffiths JF (1995) An assessment of the weather generator (WXGEN) used in the erosion/productivity impact calculator. Agric For Meteorol 73:115–133

Wilks DS, Wilby RL (1999) The weather generation game: a review of stochastic weather models. Prog Phys Geog 23:329–357

Wu GF, Liu YL, Wang TJ (2007) Methods and strategy for modeling daily global solar radiation with measured meteorological data—a case study in Nanchang station, China. Energy Convers Manag 48:2447–2452

Xia Y, Winterhalter M, Fabian P (2000) Interpolation of daily global solar radiation with thin plate smoothing splines. Theor Appl Climatol 66:109–115

Yang Y, Wang D, Lv W, Mo Y, Ding L (2009) Solar radiation standard and its values transfer system in China.www.wmo.int/pages/prog/www/IMOP/publications/IOM-96_TECO-2008/P1(52)_Yang_China.pdf

Zhou J, Wu Y, Yan G (2005) General formula for estimation of monthly average daily global solar radiation in China. Energy Convers Manag 46:257–268