Renewable Energy 36 (2011) 413–420

Contents lists available at ScienceDirect

Renewable Energy

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

Technical Note

Estimation of monthly solar radiation from measured temperatures using support vector machines – A case study

Ji-Long Chen [a,c], Hong-Bin Liu [a,c], Wei Wu [b,c,*], De-Ti Xie [a,c]

[a] College of Resources and Environment, Southwest University, Chongqing 400716, China
[b] College of Computer and Information Science, Southwest University, Chongqing 400716, China
[c] Chonging Key Laboratory of Digital Agriculture, Chongqing 400716, China

ARTICLE INFO

Article history:
Received 2 November 2009
Accepted 16 June 2010
Available online 15 July 2010

Keywords:
Input feature
Kernel function
Comparison
Empirical model
Support vector machine

ABSTRACT

Solar radiation is the principal and fundamental energy for many physical, chemical and biological processes. However, it is measured at a very limited number of meteorological stations in the world. This paper presented the methods of monthly mean daily solar radiation estimation using support vector machines (SVMs), which is a relatively new machine learning algorithm based on the statistical learning theory. The main objective of this paper was to examine the feasibility of SVMs in estimating monthly solar radiation using air temperatures. Measured long-term monthly air temperatures including maximum and minimum temperatures (T_{max} and T_{min}, respectively) were gathered and analyzed at Chongqing meteorological station, China. Seven combinations of air temperatures, namely, (1) T_{max}, (2) T_{min}, (3) $T_{max} - T_{min}$, (4) T_{max} and T_{min}, (5) T_{max} and $T_{max} - T_{min}$, (6) T_{min} and $T_{max} - T_{min}$, and (7) T_{max}, T_{min}, and $T_{max} - T_{min}$, were served as input features for SVM models. Three equations including linear, polynomial, and radial basis function were used as kernel functions. The performances were evaluated using root mean square error (RMSE), relative root mean square error (RRMSE), Nash-Sutcliffe (NSE), and determination coefficient (R^2). The developed SVM models were also compared with several empirical temperature-based models. Comparison analyses showed that the newly developed SVM model using T_{max} and T_{min} with polynomial kernel function performed better than other SVM models and empirical methods with highest NSE of 0.999, R^2 of 0.969, lowest RMSE of 0.833 MJ m^{-2} and RRMSE of 9.00%. The results showed that the SVM methodology may be a promising alternative to the traditional approaches for predicting solar radiation where the records of air temperatures are available.

1. Introduction

Solar radiation (R_s) at the earth's surface is the principal and fundamental energy for many physical, chemical and biological processes, such as crop growth, plant photosynthesis [1], and it is also an essential and important variable to many simulation models studies, such as agriculture, environment, hydrology, meteorology and ecology [2]. Hence, accurate records of R_s are of vital importance [3]. The best R_s at the specific sites should be continuously and accurately over a long time [4]. However, in many cases it is not readily available due to the cost and difficulty of maintenance and calibration of the measurement equipment, only a few meteorological stations measure R_s [2]. For example, in USA, less than 1% of meteorological stations are recording R_s. In China, more than 2000 stations have record of meteorological data, only 98 of them are measuring R_s [5]. The ratio of meteorological stations recording R_s to those recording temperature is about 1:500 around the world [6]. Moreover, even at these stations there may be many days when R_s data are missing or lie outside the expected range [7]. The lack of R_s became the major limitation to the researches and applications that require R_s as an important influencing factor [8]. This has led researchers to develop methods to estimate R_s for the sites where no direct R_s are available, or for the days when R_s are missing.

Three major methods including satellite-derived [9], stochastic algorithm [10] and empirical relationships [11–14] have been developed for this purpose. The satellite-derived method is promising for estimating R_s data over large regions, but it is relatively new and may suffer from the shortage of the historical meteorological data [7]. Stochastic weather generators are useful to generate daily simulations from data averages. However, the generated data cannot be used for model validation for a particular

* Corresponding author. College of Computer and Information Science, Chongqing Key Laboratory of Digital Agriculture, Southwest University, Chongqing 400716, China. Tel.: +86 23 6825 1069; fax: +86 23 6825 0444.
E-mail address: wuwei_star@163.com (W. Wu).

0960-1481/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.renene.2010.06.024

period of time as the data may not match the actual weather at a particular time. It fails to generate weather extremes reliably [15]. A common practice is to use empirical relationships method to estimate R_s using other readily available meteorological data, such as sunshine duration, maximum and minimum temperatures, and humidity. Previous studies have shown that R_s estimate models using sunshine duration gave better performance than that using other meteorological data [16]. However, it is often limited since sunshine duration data are absent or incomplete or inaccessible to many researchers [17]. On the contrary, air temperatures are routinely measured at most meteorological stations. Thus, in the absence of sunshine duration, lots of temperature-based models have been proposed, and numerous evaluations and modifications have subsequently been made [12–14].

Literature studies have shown that artificial neural networks (ANNs) are superior to traditional empirical models in estimating R_s [18]. Recently, support vector machines (SVMs) developed by Vapnik [19] are widely applied in computer science, bioinformatics, and environmental science [20,21]. Previous studies have proved that SVMs show better performance than neural networks and other statistical models [20]. However, there is no literature on the applications of SVM in estimating R_s with temperatures. Thus, the main objectives of this work are (1) to investigate the feasibility of SVMs to predict R_s using air temperatures, (2) to determine the impact of inputs and kernel functions on the prediction accuracy of SVMs, and (3) to evaluate the performance of the developed SVMs and empirical temperature-based models in the estimation of R_s.

2. Materials and methods

2.1. Data collection

The data set used in this work were measured at Chongqing meteorological station. The station is located at 29.58° latitude North, 106.47° longitude East, with an altitude of 416.26 m. A total of 28-year (1973–2000) monthly mean daily data were obtained from National Meteorological Information Center (NMIC), China Meteorological Administration (CMA). Air temperatures including maximum and minimum temperatures (T_{max} and T_{min}, respectively) were measured using mercury and alcohol thermometers (±0.2 °C), solar radiation is measured using pyranometer (DFY-4) (±5%), and daily values are the corresponding sums of hourly ones. The type of pyranometers used by the Chinese national measuring network was changed in 1993. However, the homogeneity of the radiation series is probably little affected, as the pyranometer models have been calibrated to the same standard following the guide of the world meteorological organization (WMO) [17,22]. Quality control tests were conducted by the suppliers. Records with missing data which were replaced by 32766 were removed from the data set. T_{max} and T_{min} were used to estimate R_s using SVMs and empirical models in the current study. Two data sets were created, about 70% (240 records, 20 years from 1973 to 1992) of the total records were used for training, and the remainder (96 records, 8 years from 1993 to 2000) for testing.

Fig. 1(a–c) showed the distributions of R_s, T_{max} and T_{min} for the study site, respectively. The R_s varied between 3.16 and 15.40 MJ m^{-2} with the maximum in August and the minimum in December. The T_{max} and T_{min} ranged from 10.22 to 33.37 °C and from 6.0 to 24.64 °C, respectively. It also can be seen that the distributions of T_{max} and T_{min} are almost similar to that of R_s but with the lowest value in January and highest in August. Pearson coefficients between T_{max} and R_s and between T_{min} and R_s were 0.91 and 0.86 ($p < 0.01$), respectively.

2.2. Theory of support vector machines

A basic theory of SVMs was presented in this section. For more detailed information on the subject, please refer to Vapnik [23] and Burges [24]. The SVM, developed by Vapnik and his coworkers, has been extensively applied in classification, regression

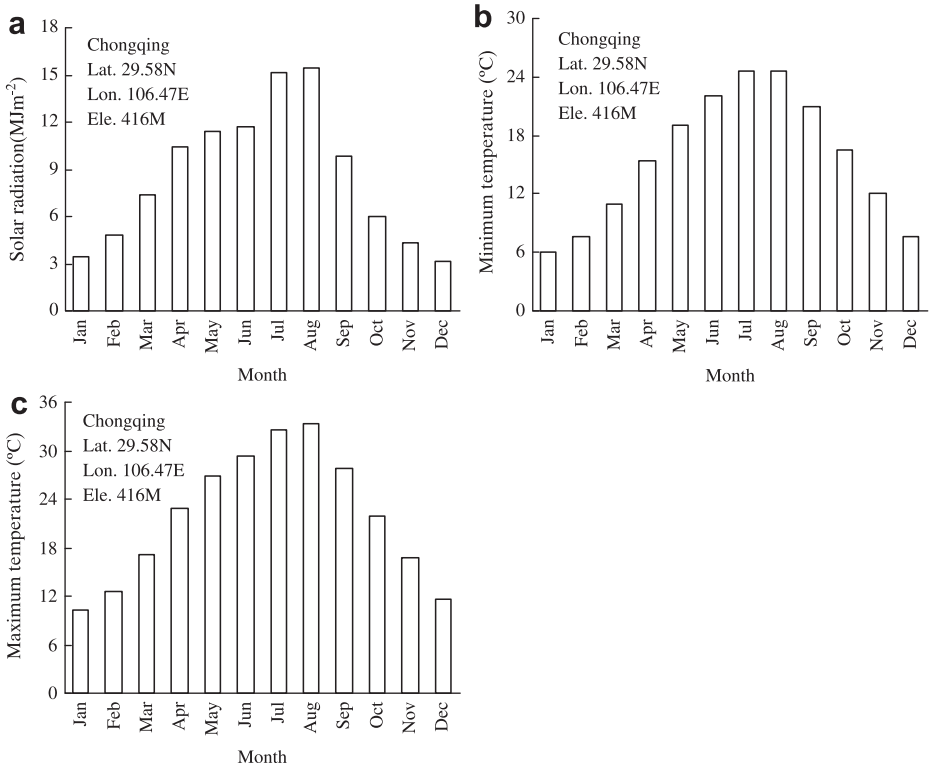

Fig. 1. Distribution of monthly mean daily solar radiation (a), minimum temperature (b), maximum temperature (c) in Chongqing.

and forecasting [25]. Comparing to traditional learning machine regressors, there are several attractive characteristics of SVMs. First of all, SVM implements the principle of Structural Risk Minimization (SRM), which attempt to minimize an upper bound of generalization error rather than minimize the local training error [19]. This is the significant difference from commonly used principle of empirical risk minimization (ERM), which employed by the traditional learning machine regressors [23]. 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 $G = \{(x_i, d_i)\}_i^n$ ($x_i$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 [23]:

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

where $\phi(x)$ 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 [19,23]:

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

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

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

The term $1/2\|w\|^2$ 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 is called 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 [19]:

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

Subjected to

$$\begin{aligned} &d_i - w\phi(x_i) - b_i \leq \varepsilon + \zeta_i \\ &w\phi(x_i) + b_i - d_i \leq \varepsilon + \zeta_i^*, \zeta_i^* \geq 0 \end{aligned} \tag{5}$$

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

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

The detail computation procedure can be found in Vapnik [25].

2.3. Kernel function

In Eq. (6), 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 x_i and x_j in the feature space $\varphi(x_i)$ and $\varphi(x_j)$, that is, $K(x_i, x_j) = \varphi(x_i) \times \varphi(x_j)$. Thus, all computations related to $\varphi(x)$ can be performed by the kernel function in input space [19]. The elegance of using the kernel function is that one can deal with feature spaces of arbitrary dimensionality without having to compute the map $\varphi(x)$ explicitly. Any function satisfying Mercer's condition can be used as kernel function [25]. The typical examples of kernel function are linear $K(x_i, x_j) = x_i \times x_j$, polynomial $K(x_i, x_j) = (x_i \times x_j + 1)^d$ and Radial basis function (RBF) $K(x_i, x_j) = \exp(-\gamma\|x_i, -x_j\|^2)$, where d and γ are the kernel parameters. The kernel parameter should be carefully chosen as it implicitly defines the structure of the high dimensional feature space $\varphi(x)$ and thus controls the complexity of the final solution.

Seven combinations of T_{max}, T_{min}, and difference between T_{max} and T_{min}, namely, (1) T_{max}, (2) T_{min}, (3) $T_{max} - T_{min}$, (4) T_{max} and T_{min}, (5) T_{max} and $T_{max} - T_{min}$, (6) T_{min} and $T_{max} - T_{min}$, and (7) T_{max}, T_{min}, and $T_{max} - T_{min}$, were served as input features for SVM models. Linear, polynomial, and RBF equations were used as kernel functions. Hence, a total of 21 SVM models were explored in the present study (Table 1).

2.4. Temperature-based models

2.4.1. H–S model

Hargreaves and Samani [12] made an initial study on using T_{max} and T_{min} to estimate solar radiation by the following equation:

$$R_s = K_r\sqrt{(T_{max} - T_{min})}R_a \tag{7}$$

where R_a is extraterrestrial radiation, which is a function of latitude and day of the year, the detailed procedure of the calculation of R_a can be found in Allen et al.[26]. K_r is an empirical coefficient, which was recommended to be 0.16 for interior regions and 0.19 for coastal regions. In this work, K_r (0.1052) was locally determined using the training data set.

2.4.2. H–S–A model

Annandale et al. [14] modified the H–S model by introducing a correction factor and obtained the following equation:

$$R_s = K_r\left(1 + 2.7 \times 10^{-5}Z\right)\sqrt{(T_{max} - T_{min})}R_a \tag{8}$$

where Z is the elevation in m.

Table 1
The developed SVM models using different inputs with different kernel functions.

Model	Kernel function	Input attribute[a]
SVM_{l1}	Linear	T_{max}
SVM_{l2}	Linear	T_{min}
SVM_{l3}	Linear	$T_{max} - T_{min}$
SVM_{l4}	Linear	T_{max}, T_{min}
SVM_{l5}	Linear	T_{max}, $T_{max} - T_{min}$
SVM_{l6}	Linear	T_{min}, $T_{max} - T_{min}$
SVM_{l7}	Linear	T_{max}, T_{min}, $T_{max} - T_{min}$
SVM_{p1}	Polynomial	T_{max}
SVM_{p2}	Polynomial	T_{min}
SVM_{p3}	Polynomial	$T_{max} - T_{min}$
SVM_{p4}	Polynomial	T_{max}, T_{min}
SVM_{p5}	Polynomial	T_{max}, $T_{max} - T_{min}$
SVM_{p6}	Polynomial	T_{min}, $T_{max} - T_{min}$
SVM_{p7}	Polynomial	T_{max}, T_{min}, $T_{max} - T_{min}$
SVM_{r1}	Radial Base Function	T_{max}
SVM_{r2}	Radial Base Function	T_{min}
SVM_{r3}	Radial Base Function	$T_{max} - T_{min}$
SVM_{r4}	Radial Base Function	T_{max}, T_{min}
SVM_{r5}	Radial Base Function	T_{max}, $T_{max} - T_{min}$
SVM_{r6}	Radial Base Function	T_{min}, $T_{max} - T_{min}$
SVM_{r7}	Radial Base Function	T_{max}, T_{min}, $T_{max} - T_{min}$

[a] T_{max}, T_{min}, $T_{max} - T_{min}$ are monthly mean daily maximum and minimum temperatures, difference between monthly mean daily maximum and minimum temperatures, respectively.

Table 2
Root mean square error (RMSE in MJ m^{-2}), relative root mean square error (RRMSE) and Nash-Sutcliffe coefficients (NSEs) of the developed SVM models.

Model[a]	RMSE	RRMSE	NSE	Model[a]	RMSE	RRMSE	NSE	Model[a]	RMSE	RRMSE	NSE	Overall		
												RMSE	RRMSE	NSE
SVM_{l1}	1.5387	19.43%	0.8828	SVM_{p1}	1.3528	18.90%	0.9984	SVM_{r1}	1.2907	18.23%	0.9985	1.3941	18.85%	0.9599
SVM_{l2}	1.8496	24.00%	0.8307	SVM_{p2}	1.7626	23.00%	0.9979	SVM_{r2}	1.7115	24.26%	0.9979	1.7746	23.75%	0.9422
SVM_{l3}	1.2924	15.65%	0.9173	SVM_{p3}	1.2978	11.18%	0.9984	SVM_{r3}	1.2954	12.37%	0.9985	1.2952	13.07%	0.9714
SVM_{l4}	1.0757	10.11%	0.9427	SVM_{p4}	0.8330	9.00%	0.9990	SVM_{r4}	0.8704	9.43%	0.9989	0.9264	9.51%	0.9802
SVM_{l5}	1.1377	11.94%	0.9359	SVM_{p5}	0.921	10.50%	0.9989	SVM_{r5}	1.1859	12.04%	0.9986	1.0815	11.49%	0.9778
SVM_{l6}	1.1461	12.17%	0.9350	SVM_{p6}	0.9413	9.91%	0.9989	SVM_{r6}	1.3867	11.23%	0.9983	1.1580	11.10%	0.9774
SVM_{l7}	1.0918	10.64%	0.9410	SVM_{p7}	0.9041	9.82%	0.9989	SVM_{r7}	1.1249	11.21%	0.9987	1.0403	10.56%	0.9795
Overall														
	1.3045	14.85%	0.9122		1.1447	13.19%	0.9986		1.2665	14.11%	0.9985			

[a] See Table 1.

2.4.3. B–C model

Bristow and Campbell [13] studied the relationship between R_s and T_{max}, T_{min} and developed the equation given below:

$$R_s = AR_a\left[1 - \exp\left(-B(\Delta T)_i\right)\right] \tag{9}$$

where A, B, and C are empirical coefficients, values widely used for these coefficients are 0.7 for A, 2.4 for C and the value of B can be calculated from the equation [10]:

$$B = 0.036\exp(-0.154\overline{\Delta T}) \tag{10}$$

where $\overline{\Delta T}$ is the monthly average of the ΔT_i daily values, ΔT_i is the air temperature difference calculated by subtracting the average minimum temperature of the current and the next day from the maximum temperature of the current day:

$$\Delta T_i = T_{max,i} - \frac{T_{min,i} + T_{min,i+1}}{2} \tag{11}$$

where $T_{max,i}$, $T_{min,i}$, $T_{min,i+1}$ are the maximum, minimum temperatures in ith day, minimum temperature in $(i+1)$th day, respectively [27].

2.4.4. Chen model

Chen et al. [5] validated three air temperature-based models in China and presented a new logarithmic expression:

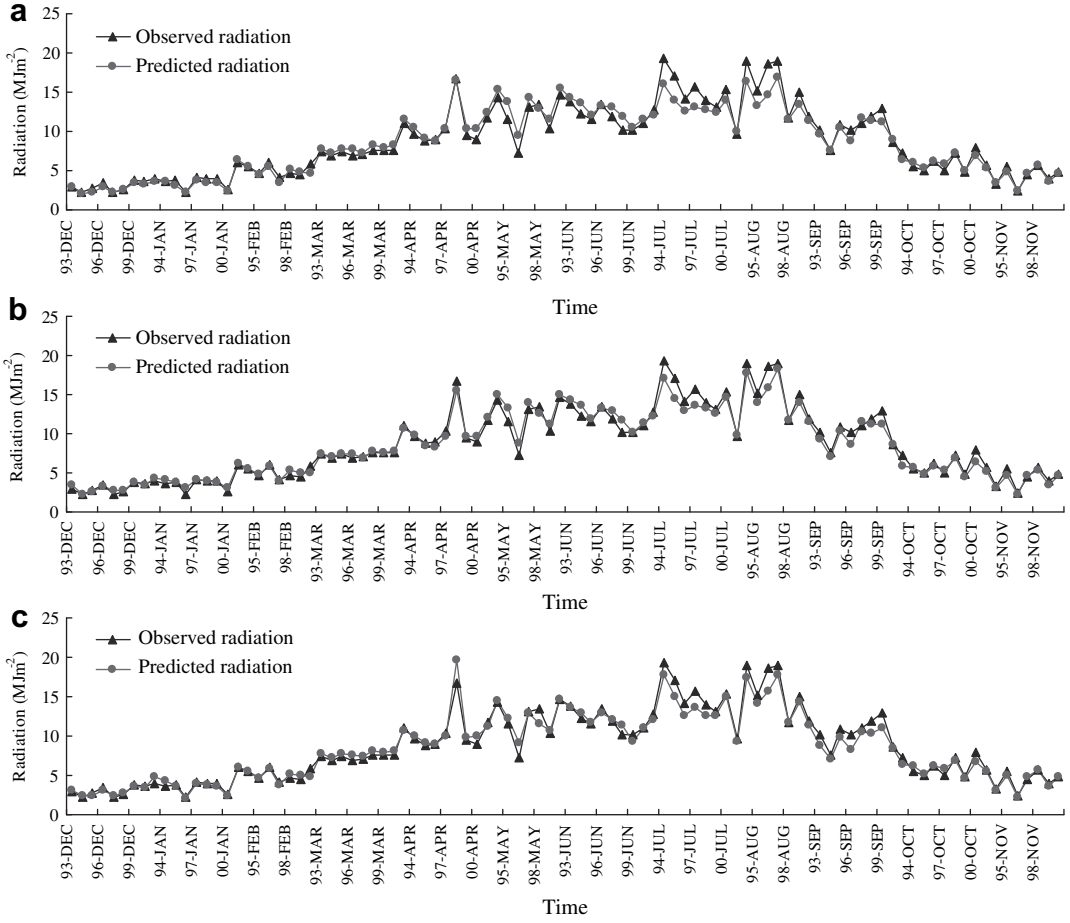

Fig. 2. Plots of the observed and estimated radiation by SVM models using maximum and minimum temperatures with linear (a), polynomial (b) and radial basis function (c) kernel functions.

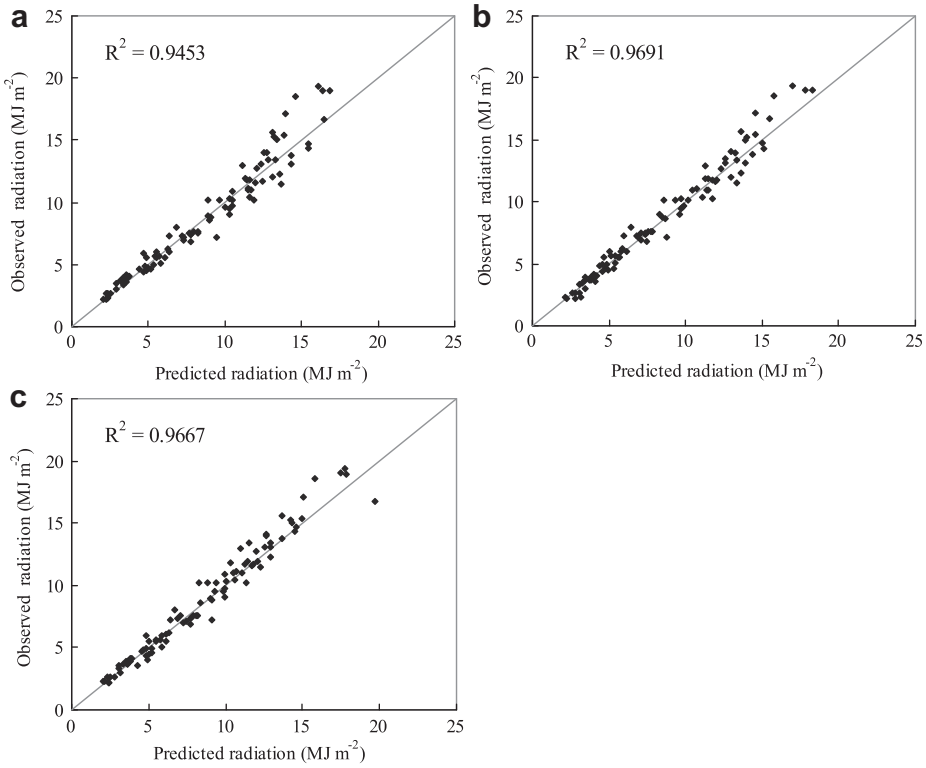

Fig. 3. Scatter plots of the observed vs. predicted radiation by the SVM models using maximum and minimum temperatures with linear (a), polynomial (b) and radial basis function (c) kernel functions.

$$R_s = AR_a\ln(T_{max} - T_{min}) + B \tag{12}$$

where A and B are empirical coefficients.

2.4.5. Local regression models

Except these well-known models, two equations were created using the training data set on the basis of the relationship between R_s and temperatures of Chongqing (hereafter local model 1 (Eq. (13)) and local model 2 (Eq. (14)), respectively).

$$R_s = R_a\left(0.225\sqrt{T_{max} - T_{min}} - 0.3047\right) \tag{13}$$

$$R_s = R_a(0.0461(T_{max} - T_{min}) - 0.0367) \tag{14}$$

2.4.6. Performance criteria

To assess the performance of these models, root mean square error (RMSE), relative root mean square error (RRMSE), Nash-Sutcliffe coefficient (NSE), mean absolute error (MAE), relative mean absolute error (RMAE) and determination coefficient (R^2) were determined. The metric R^2 was adopted to measure the correlation of the observed and predicted values. The former five indicators were calculated by the following equations.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}\left(y_i - \widehat{y}_i\right)^2}{n}} \tag{15}$$

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

$$\mathrm{NSE} = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \widehat{y}_i\right)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \tag{17}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left(\left|y_i - \widehat{y}_i\right|\right) \tag{18}$$

$$\mathrm{RMAE} = \frac{1}{n}\sum_{i=1}^{n}\left(\left|\frac{y_i - \widehat{y}_i}{\widehat{y}_i}\right|\right) \tag{19}$$

Table 3
Root mean square error (RMSE in MJ m^{-2}), Relative root mean square error (RRMSE), Nash-Sutcliffe coefficients (NSEs), Mean absolute error (MAE in MJ m^{-2}), Relative mean absolute error (RMAE) and relative improvement (RI) of SVM_{p4} and conventional regression methods.

Model	RMSE	RRMSE	RI_{rmse}	NSE	RI_{nse}	MAE	RMAE
SVM_{p4}[a]	0.8330	9.00%		0.9990		0.5932	7.01%
H–S[b]	1.9624	24.77%	57.55%	0.8094	23.43%	1.4253	19.25%
H–S–A[c]	1.9439	25.19%	57.15%	0.8129	22.88%	1.4291	19.59%
B–C[d]	1.5089	24.85%	44.80%	0.8873	12.59%	1.2075	18.72%
Chen[e]	1.3693	16.06%	39.16%	0.9072	10.12%	0.9981	12.34%
Local model 1[f]	1.1581	11.86%	28.07%	0.9336	7.00%	0.8387	9.11%
Local model 2[g]	1.2385	12.99%	32.74%	0.9241	8.11%	0.9059	10.60%

[a] See Table 1.
[b] Hargreaves and Samani. [12].
[c] Annandale et al. [14].
[d] Bristow and Campbell [13].
[e] Chen et al. [5].
[f] local regressive model (Eq. (13)).
[g] local regressive model (Eq. (14)).

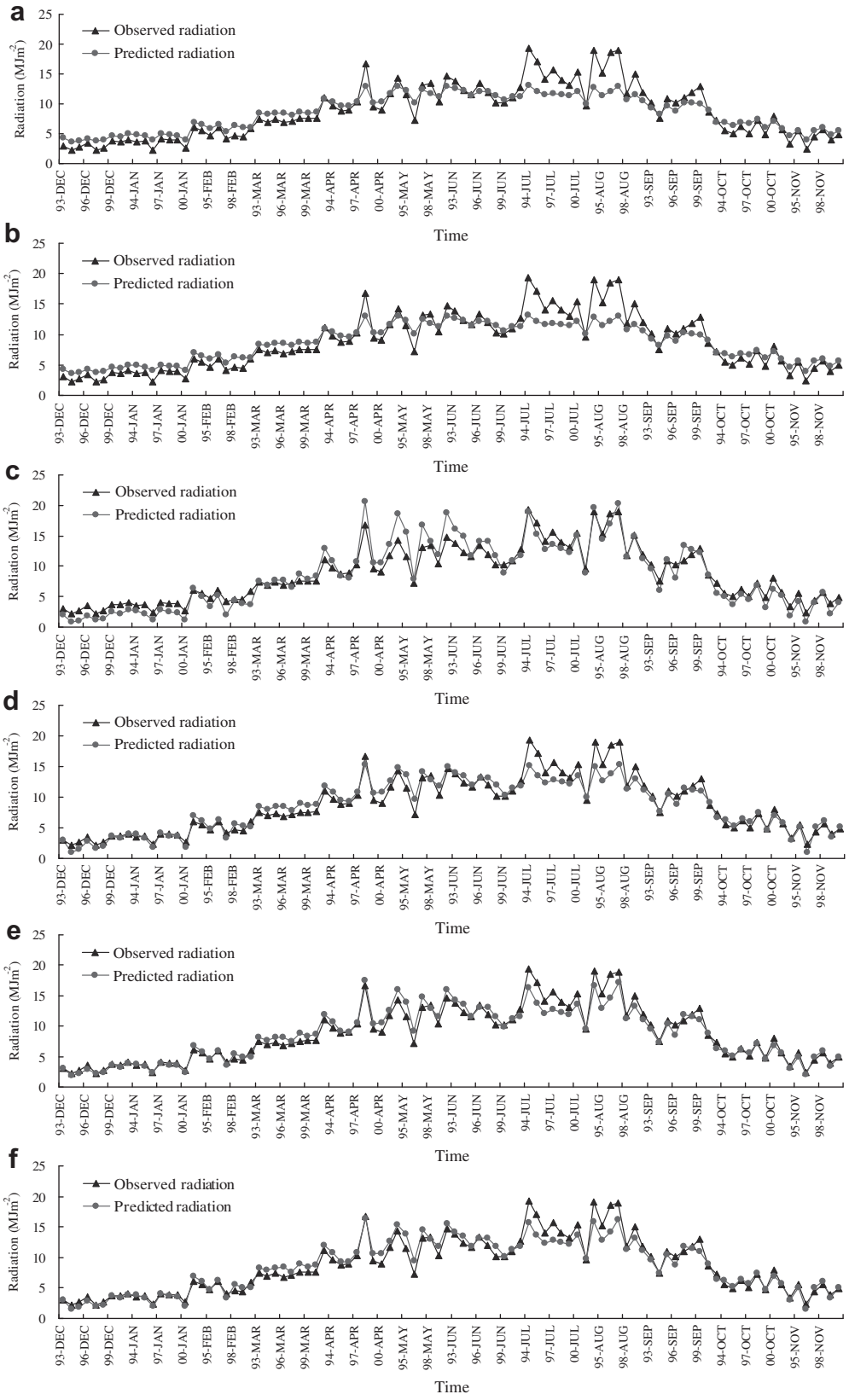

Fig. 4. Plots of the observed and estimated radiation by H–S (a), H–S–A (b), B–C (c), Chen (d), Local model 1 (e), and Local model 2 (f).

where n, y, $\hat{y}$ and $\bar{y}$ represent the number of testing data, the observation, the predicted value and the average value of the observation, respectively. Lower values of RMSE, RRMSE, MAE and RMAE indicate a better estimation of a model. A model is more efficient when NSE is closer to 1.

The relative improvement of RMSE (RI_{rmse}) and NSE (RI_{nse}) were used to measure the improvement on the prediction accuracy of SVM over the reference methods (temperature-based models):

$$RI_{\text{rmse}} = \frac{\text{RMSE}_R - \text{RMSE}_E}{\text{RMSE}_R} \tag{20}$$

$$RI_{\text{nse}} = \frac{\text{NSE}_E - \text{NSE}_R}{\text{NSE}_R} \tag{21}$$

where RMSE_R and NSE_R are the root mean square error and Nash-Sutcliffe coefficient for the reference methods, respectively. RMSE_E and NSE_E are the root mean square error and Nash-Sutcliffe coefficient for SVM models, respectively.

3. Results and discussion

3.1. Performance of SVM models

Table 2 displayed the performance indicators of the developed SVM models. Overall, all the SVM models gave good prediction performance with RMSEs < 1.85 MJ m^{-2}, RRMSEs $< 25\%$ and NSEs > 0.83. It was evident that the model using $T_{\max}$ and $T_{\min}$ with polynomial (SVM_{p4}) gave the best performance with the lowest RMSE of 0.833 MJ m^{-2}, RRMSE of 9.00%, highest NSE of 0.999 and R^2 of 0.9691 (Table 2 and Fig. 3b).

Obviously, both kernel functions and inputs have influence on the accuracy of SVM models. On average, SVMs with polynomial equation gave the best performance with lowest RMSE (1.1447 MJ m^{-2}) and RRMSE (13.19%) and highest NSE (0.9986) followed by SVMs with RBF (RMSE = 1.2665 MJ m^{-2}, RRMSE of 14.11%, NSE = 0.9985) and linear (RMSE = 1.3045 MJ m^{-2}, RRMSE of 14.85%, NSE = 0.9122). However, in many SVM researches RBF has been commonly used and recommended other than polynomial kernel function [20,21]. It is probable that polynomial kernel

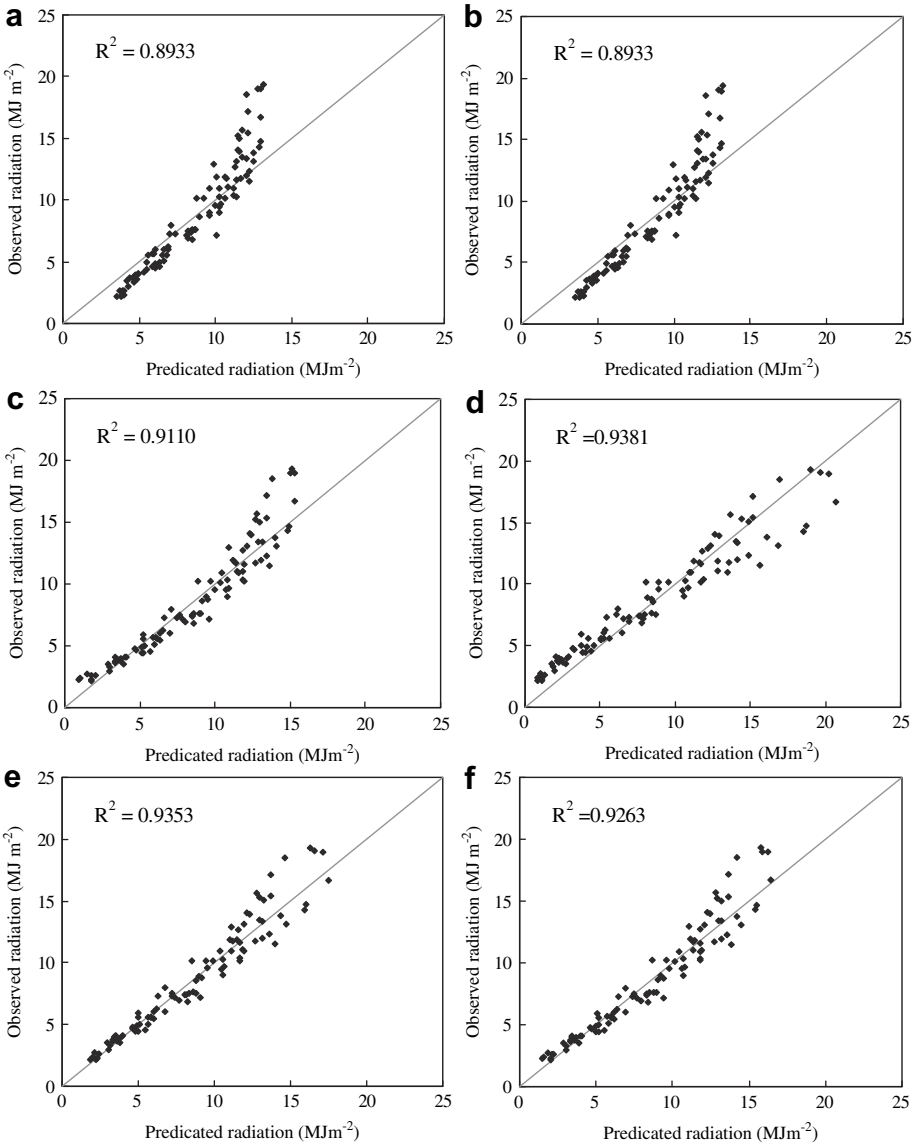

Fig. 5. Scatter plots of the observed vs. predicted radiation by H–S (a), H–S–A(b), B–C (c), Chen (d), Local model 1 (e), and Local model 2 (f).

function has more hyperparameters, which influence the complexity of the computations [21]. The results of this study indicated that selecting an appropriate kernel function is of importance for the accuracy of an SVM model.

The averaged RMSE of SVMs with different input attributes was in order: T_{max}, $T_{min} < T_{max}$, T_{min}, $T_{max} - T_{min} < T_{max}$, $T_{max} - T_{min} < T_{min}$, $T_{max} - T_{min} < T_{max} - T_{min} < T_{max} < T_{min}$. Averaged RRMSE of SVMs with different input attributes was in order: T_{max}, $T_{min} < T_{max}$, T_{min}, $T_{max} - T_{min} < T_{min}$, $T_{max} - T_{min} < T_{max}$, $T_{max} - T_{min} < T_{max} - T_{min} < T_{max} < T_{min}$. Averaged NSE of SVMs with different input attributes was in order: T_{max}, $T_{min} > T_{max}$, T_{min}, $T_{max} - T_{min} > T_{max}$, $T_{max} - T_{min} > T_{min}$, $T_{max} - T_{min} > T_{max} - T_{min} > T_{max} > T_{min}$. Models with linear, polynomial and RBF functions using combined inputs gave an averaged RMSE of 1.113 MJ m^{-2}, RRMSE of 11.21% and NSE of 0.939, RMSE of 0.899 MJ m^{-2}, RRMSE of 9.81% and NSE of 0.999, and RMSE of 1.142, RRMSE of 10.98% and NSE of 0.999, respectively.

Comparison within the models with same kernel function showed that models using T_{max} and T_{min}, namely, SVM_{l4}, SVM_{p4}, and SVM_{r4}, performed better with values of RMSE of 1.076 MJ m^{-2}, 0.833 MJ m^{-2}, and 0.870 MJ m^{-2}, RRMSE of 10.11%, 9.00%, 9.43% and NSE of 0.943, 0.999, and 0.999, respectively (Table 2). The predicted R_s by SVM_{l4}, SVM_{p4}, and SVM_{r4} were close to the observed values (Fig. 2), indicating that the three models showed good generalization. The higher R^2 (0.9453, 0.9691, 0.9667 for SVM_{l4}, SVM_{p4}, and SVM_{r4}, respectively) further confirmed the good performance of SVM_{l4}, SVM_{p4}, and SVM_{r4} (Fig. 3).

3.2. Comparison of SVM between temperature-based models

The performance of the best SVM model (SVM_{p4}) and empirical temperature-based models were presented in Table 3 and Figs. 4 and 5. Comparison analyses showed that SVM_{p4} gave lowest RMSE of 0.833 MJ m^{-2} and RRMSE of 9.00% and highest NSE of 0.999 and the R^2 of 0.9691 (Table 3, Figs. 3b and 5). The prediction by SVM_{p4} was much closer to the observation than the empirical models (Figs. 2b and 4), with the lowest MAE of 0.593 MJ m^{-2} and RMAE of 7.01%(Table 3). In terms of the RMSE, the relative improvement of SVM_{p4} ranged from 28.07% on Local model 1 to 57.55% on H–S model, with an averaged of 45.56%(Table 3). In terms of the NSE, the relative improvement of SVM_{p4} ranged from 7% on Local model 1 to 23.43% on H–S model, with an averaged of 13.64% (Table 3). These results proved that SVM_{p4} outperformed the empirical temperature-based models in this study site.

4. Conclusion

Solar radiation is an essential and important variable to many simulation models, thus, it is urgently required to develop a novel and more accurate method to estimate solar radiation when it is not readily available. The objective of this work was to investigate the feasibility of SVM model to estimate solar radiation with routinely measured temperature. In this work, different kernel functions and inputs for the SVM models were used. The performances of the SVM models were evaluated by several statistical indices. These models were also compared with several empirical temperature-based equations. Both kernel function and input attributes have influence on the accuracy of SVM models. Comparison analyses showed that SVM using T_{max} and T_{min} with polynomial kernel function outperformed other models. Future works will focus on the predictability of SVM methodology on a regional scale.

Acknowledgments

The work was supported by China Postdoctoral Science Foundation (No. 20090450794) and Chongqing Key Laboratory of Digital Agriculture (China).

References

[1] Meza F, Varas E. Estimation of mean monthly solar global radiation as a function of temperature. Agr For Meteorol 2000;100:231–41.
[2] Hunt LA, Kuchar L, Swanton CJ. Estimation of solar radiation for use in crop modelling. Agr For Meteorol 1998;91:293–300.
[3] El-Sebaii AA, Trabea AA. Estimation of horizontal diffuse solar radiation in Egypt. Energ Convers Manage 2003;44:2471–82.
[4] Wu GF, Liu YL, Wang TJ. Methods and strategy for modeling daily global solar radiation with measured meteorological data – a case study in Nanchang station, China. Energ Convers Manage 2007;48:2447–52.
[5] Chen RS, Ersi K, Yang JP, Lu SH, Zhao WZ. Validation of five global radiation models with measured daily data in China. Energ Convers Manage 2004;45:1759–69.
[6] NCDC (National Climatic Data Center). Cooperative summary of the day, dataset TD 3200. U.S. Department of Commerce. Asheville, NC: National Oceanographic and Atmospheric Administration, National Climatic Data Center; 1995.
[7] Abraha MG, Savage MJ. Comparison of estimates of daily solar radiation from air temperature range for application in crop simulations. Agr For Meteorol 2008;148:401–16.
[8] Hook JE, McClendon RW. Estimation of solar radiation data from long-term meteorological records. Agron J 1992;84:739–42.
[9] Pinker RT, Frouin R, Li Z. A review of satellite methods to derive shortwave irradiance. Remote Sens Environ 1995;51:108–24.
[10] Richardson CW. Stochastic simulation of daily precipitation, temperature, and solar radiation. Water Resour Res 1981;17:182–90.
[11] Ångström A. Solar and terrestrial radiation. Q J Roy Meteor Soc 1924;50:121–6.
[12] Hargreaves GH, Samani ZA. Estimating potential evapotranspiration. J Irrg Drain Engrg ASCE 1982;108:225–30.
[13] Bristow KL, Campbell GS. On the relationship between incoming solar radiation and daily maximum and minimum temperature. Agr For Meteorol 1984;31:159–66.
[14] Annandale JG, Jovanic NZ, Benade N, Allen RG. Software for missing data error analysis of Penman-Monteith reference evapotranspiration. Irrig Sci 2002;21:57–67.
[15] Wallis TWR, Griffiths JF. An assessment of the weather generator (WXGEN) used in the erosion/productivity impact calculator. Agr For Meteorol 1995;73:115–33.
[16] Trnka M, Žalud Z, Eitzinger J, Dubrovský M. Global solar radiation in Central European lowlands estimated by various empirical formulae. Agr For Meteorol 2005;131:54–76.
[17] Liu XY, Mei XR, Li YZ, Wang QS, Jensen JR, Zhang XQ, et al. Evaluation of temperature-based global solar radiation models in China. Agr For Meteorol 2009;149:1433–46.
[18] Al-Alawi SM, Al-Hinai HA. An ANN-based approach for predicting global solar radiation in locations with no measurements. Renewable Energy 1998;14:199–204.
[19] Vapnik V. The nature of statistical learning theory. New York: Springer; 1995.
[20] Liu HB, Xie DT, Wu W. Soil water content forecasting by ANN and SVM hybrid architecture. Environ Monit Assess 2008;143:187–93.
[21] Dong B, Cao C, Lee SE. Applying support vector machines to predict building energy consumption in tropical region. Energy and Buildings 2005;37:545–53.
[22] Yang Y, Wang D, Lv W, Mo Y, Ding L. Solar radiation standard and its values transfer system in China, www.wmo.int/pages/prog/www/IMOP/publications/IOM-96_ECO-2008/P1(52)_Yang_China.pdf.
[23] Vapnik V. Statistical learning theory. New York: Wiley; 1998.
[24] Burges CJC. A tutorial on support vector machines for pattern recognition. Data Min Knowl Disc 1998;2(2):121–67.
[25] Vapnik V, Golowich SE, Smola AJ. Support vector method for function approximation, regression estimation and signal processing. Adv Neural Inf Process Syst 1996;9:281–7.
[26] Allen RG, Pereira LS, Raes D, Smith M. Crop evapotranspiration – guidelines for computing crop water requirements – FAO irrigation and drainage paper 56. Rome: Food and Agriculture Organization of the United Nations; 1998.
[27] Liu DL, Scott BJ. Estimation of solar radiation in Australia from rainfall and temperature observations. Agr For Meteorol 2001;106:41–59.