sustainability

MDPI

*Article*

# Improvement in Solar-Radiation Forecasting Based on Evolutionary KNEA Method and Numerical Weather Prediction

**Guosheng Duan [1], Lifeng Wu [2,3,*], Fa Liu [4], Yicheng Wang [2] and Shaofei Wu [3]**

[1] Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China; guoshengduan@bnu.edu.cn
[2] State Key Laboratory of Simulation and Regulation of Water Cycle in River Basin, China Institute of Water Resources and Hydropower Research, Beijing 100038, China; wangych@iwhr.com
[3] School of Hydraulic and Ecological Engineering, Nanchang Institute of Technology, Nanchang 330099, China; sfw17@nit.edu.cn
[4] Key Laboratory of Water Cycle and Related Land Surface Processes, Institute of Geographic Science and Science and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China; liufa@igsnrr.ac.cn
* Correspondence: lifengwu@nit.edu.cn

**Abstract:** Accurate forecasting of solar radiation (Rs) is significant to photovoltaic power generation and agricultural management. The National Centers for Environmental Prediction (NECP) has released its latest Global Ensemble Forecast System version 12 (GEFSv12) prediction product; however, the capability of this numerical weather product for Rs forecasting has not been evaluated. This study intends to establish a coupling algorithm based on a bat algorithm (BA) and Kernel-based nonlinear extension of Arps decline (KNEA) for post-processing 1–3 d ahead Rs forecasting based on the GEFSv12 in Xinjiang of China. The new model also compares two empirical statistical methods, which were quantile mapping (QM) and Equiratio cumulative distribution function matching (EDCDFm), and compares six machine-learning methods, e.g., long-short term memory (LSTM), support vector machine (SVM), XGBoost, KNEA, BA-SVM, BA-XGBoost. The results show that the accuracy of forecasting Rs from all of the models decreases with the extension of the forecast period. Compared with the GEFS raw Rs data over the four stations, the RMSE and MAE of QM and EDCDFm models decreased by 20% and 15%, respectively. In addition, the BA-KNEA model was superior to the GEFSv12 raw Rs data and other post-processing methods, with $R^2$ = 0.782–0.829, RMSE = 3.240–3.685 MJ $m^{-2}$ $d^{-1}$, MAE = 2.465–2.799 MJ $m^{-2}$ $d^{-1}$, and NRMSE = 0.152–0.173.

**Keywords:** forecasting; solar radiation; Global Ensemble Forecast System; bat algorithm

check for updates

**Citation:** Duan, G.; Wu, L.; Liu, F.; Wang, Y.; Wu, S. Improvement in Solar-Radiation Forecasting Based on Evolutionary KNEA Method and Numerical Weather Prediction. *Sustainability* **2022**, *14*, 6824. https://doi.org/10.3390/su14116824

Academic Editors: Zhengxin Wang, Song Ding, Xin Ma and Wendong Yang

Received: 26 April 2022
Accepted: 30 May 2022
Published: 2 June 2022



## 1. Introduction

Solar radiation is the primary source of surface energy, which drives carbon and water exchanges between the atmosphere and terrestrial ecosystems [1]. Population growth, limited fossil fuels, and environmental pollution have caused the rapid development of renewable energy sources such as solar and wind power. However, in many solar energy applications, accurate information about the presence of solar energy is required [2]. Solar measuring equipment is much more expensive than other meteorological parameters such as temperature, relative humidity and wind speed. More than 2400 weather stations in China record meteorological data, while only about 5% of stations observe global solar radiation (Rs). Therefore, models need to be developed for stations with no solar-radiation records, to estimate solar radiation [3]. Three main methods are used to calculate daily global solar radiation, i.e., satellite-derived, stochastic and meteorological-based processes [4]. The satellite-derived method can receive the reflectivity of the Earth's atmosphere of the irradiation, invert the daily radiation value and estimate the solar radiation in a large area.

Nevertheless, the uncertainty of satellite-based solar-radiation remote sensing can be high in cloudy and polluted areas. Stochastic algorithms depend on history; a statistical summary of radiation information is used to infer the probability of future radiation, which requires the support of existing high-quality historical radiation-observation data. Weather-based approaches aim to establish relationships between solar radiation and other, more readily available, meteorological elements. This method is by far the most widely used.

Recently, machine-learning models, due to their super nonlinear fitting ability, have been widely used in the simulation of natural phenomena, agriculture, engineering and the economy, also including Rs predicting/forecasting. Rehman and Mohandes [5] used an artificial neural network (ANN) to estimate solar radiation in Abha of Saudi Arabia. They found an ANN model with air temperature and relative humidity as inputs can capably estimate Rs. Quej et al. [6] assessed three approaches (SVM, ANN and ANFIS) to predict daily Rs in Yucatán, México. They declared that SVM models performed well in warm sub-humid regions. Ghimire et al. [7] explored the feasibility of using numerical weather prediction to forecast Rs. Deo et al. [8] used geo-temporal and satellite images as input data to feed the ELM method to develop an Rs model in Australia. The results show that the ELM model outperformed RF, M5T and MARS methods. Hassan et al. [9] evaluated the ability of four ML algorithms (MLP, ANFIS, SVM and RT) in modeling Rs. Based on these algorithms, sunshine-, temperature-, meteorological parameters- and day-number-based models were examined in Egypt. They verified that the MLP algorithm excelled in comparison to other models. On the other hand, many studies also show that ML is not always better, for example, as it has less precision than the dependency model [10]. Mohammadi et al. [11] compared the performance of an SVM model and ANFIS in predicting Rs under temperature data only with the data of Iran. It was found that the SVM model using an RBF kernel function had the highest accuracy. Feng et al. [12] used six machine-learning models to map daily global solar radiation and photovoltaic power in the Loess Plateau of China. In addition, the prediction of Rs by kernel-based machine-learning models has been widely reported in northwest China [12], humid regions of China [13], air-pollution regions of north China [14,15], Algeria [16], Spain [17], other regions around the world [18], also including diffuse radiation [19]. The kernel-based model also been used to map the solar photovoltaic potential of China [20,21].

Recently, deep-learning models have been gradually applied to the prediction of solar radiation, including LSTM algorithms, which are good at mining time-series information [22–24], and spatial processing information [25]. In addition, ML models can also be used to identify the most significant input parameters to better understand the relationship between common meteorological factors and Rs.

Voyant et al. [26] reviewed different machine-learning technologies used for solar-radiation forecasting. They pointed out that methods such as ANN and SVM were primarily used in the early stage, while methods such as regression tree and boosting tree have been used more recently. Compared with ANN, SVM, ANFIS, and decision-making, the most significant advantage tree-based methods have is processing larger data sets faster [9]. Sun et al. [27] applied an RF method to estimate Rs in an air-pollution environment. Ibrahim and Khatib [28] coupled an RF model with FFA to predict radiation on an hourly scale. Prasad et al. [29] designed a new approach named the EEMD-AOC-RF method for Rs forecasting. Firstly, this method decomposed the time lagging (t-1) data into signal data and noise data by EEMD; the data was brought into the RF model and optimized by AOC algorithm. Wu et al. [13] compared six machine-learning models (M5T, KNEA, MLP, CatBoost, RF and MARS) for predicting Rs in a sub-humid region in China. They found that the KNEA model had the highest accuracy, MLP model had the best stability, and CatBoost model had the fastest speed.

Recently, The National Centers for Environmental Prediction (NECP) released its new product, Global Ensemble Forecast System version 12 (GEFSv12) [30]. This product has up to 35 days ahead of Rs forecast data, however, its accuracy has not been evaluated. A new model-based bat algorithm and KNEA was used to forecasting Rs, and the input data was from the GEFSv12 output for the 1–3 d ahead. Therefore, the objectives of this study were: (1) to evaluate the 1–3 d ahead solar-radiation-prediction performance of GEFSv12 at four stations in northwest China; (2) to build a coupling model based on the bat algorithm and KNEA (BA–KNEA) model; and (3) compare the newly developed BA-KNEA model with the traditional empirical model and five other machine-learning models.

## 2. Materials and Methods

### *2.1. Study Region*

This study uses observational data from four radiation stations in Xinjiang of China, whose geographical locations are shown in Figure 1. The region is rich in solar-radiation resources, with an annual average 5200–6400 MJ $m^{-2}$ $y^{-1}$. The annual average air temperature is 9 °C and annual precipitation is less than 200 mm $y^{-1}$. These stations are affiliated with the Meteorological Data Center of the China Meteorological Administration, and the data include the total daily surface radiation from 2006 to 2015. The data was divided into two parts, the first part (2005–2010) was used for training the model and the other was used to test the model. When the Rs of a day was higher than the extraterrestrial radiation, the data of that day were deleted [31]. The global solar radiation for different months at each station is outlined in Table 1.

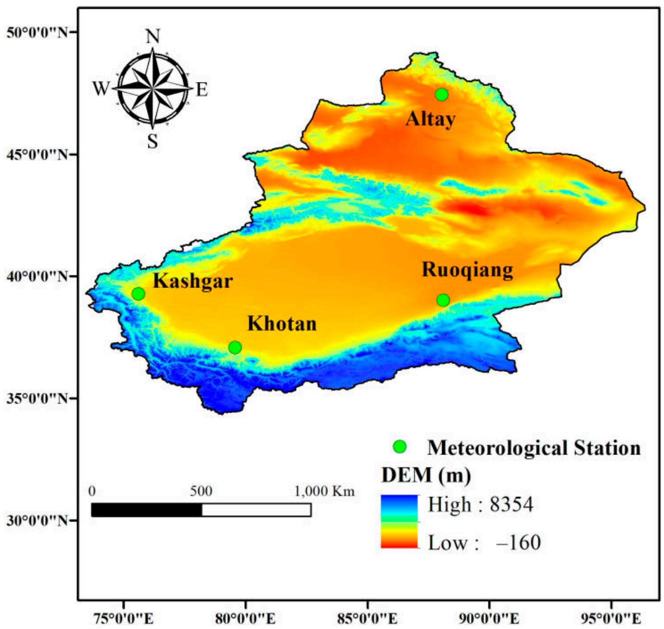


**Figure 1.** Location of the meteorological stations.

NCEP implemented its next Global Ensemble Forecasting System (GEFSv12) in summer 2020. This model upgrade, based on a deterministic and ensemble prediction system, is very different from the previous upgrade. In the NCEP operation model, a new dynamic core (FV3) is used for the first time to replace the previous spectral dynamic core [32]. The previous three categories of Zhao–Carr microphysics schemes have also been replaced by the more advanced six categories of GFDL microphysics schemes. From the perspective of the ensemble model, GEFSv12 extends the prediction period to 35 days. To better represent the considerable uncertainty related to this time scale, random physical-disturbance trends and random kinetic-energy-backscattering stochastic schemes replaced the original random general-disturbance-trend stochastic scheme, which is also a significant upgrade of the system [33]. Its spatial resolution is 25 km and its temporal resolution is 3 h. In this study, we used grid data from the mean of the four grid points around the site, including forecasting solar radiation ($Rs_f$), maximum temperature ($Tmax_f$), minimum temperature ($Tmin_f$), relative humidity ($RH_f$) at 2 m height and wind speed ($U_f$) at 10 m height every 3 h for the next 72 h, and converted the 3 h time-resolution data into daily data. That means that, for 3 h to 24 h (27 h to 48 h and 51 to 72 h), the eight data points were converted to daily scale. $Tmax_f$ and $Tmin_f$ are the highest and lowest temperature of the eight time scale in one day. $RH_f$ and $U_f$ are the mean of the eight-point time scale in one day. $Rs_f$ is the sum of the eight-point time scale in a day. The output of the models is the measured Rs corresponding to the GEFS data on the same day.

**Table 1.** Global solar radiation in different months of stations in this study.

| Station | Period | Jan. | Feb. | Mar. | Apr. | May | Jun. | Jul. | Aug. | Sept. | Oct. | Nov. | Dec. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Altay | Train | 7 ± 2.7 | 11.1 ± 3.5 | 15.7 ± 4.8 | 20.3 ± 5.9 | 23.8 ± 7.5 | 25.3 ± 7.1 | 24.2 ± 7 | 21.3 ± 6.3 | 17 ± 5.7 | 10.2 ± 4.6 | 6.1 ± 3.1 | 5.3 ± 2.4 |
| | Test | 6 ± 2.7 | 9.6 ± 3.6 | 15 ± 4.4 | 18.7 ± 5.9 | 22.5 ± 6.7 | 24.6 ± 5.6 | 23.7 ± 5.5 | 20.4 ± 5.1 | 15.9 ± 4.4 | 10 ± 4 | 6.1 ± 2.7 | 4.9 ± 2.3 |
| Kashgar | Train | 8.3 ± 2.5 | 9.6 ± 3.8 | 13.8 ± 4.6 | 19 ± 5.8 | 22.3 ± 6.2 | 26.4 ± 5 | 25.2 ± 4.6 | 21.3 ± 5 | 17.3 ± 4.4 | 13.1 ± 3.2 | 8.3 ± 2.4 | 6.2 ± 1.9 |
| | Test | 6.8 ± 2.4 | 9.1 ± 3.5 | 13 ± 4.7 | 17.2 ± 5.6 | 20.7 ± 6.1 | 24.7 ± 5 | 22.9 ± 5.6 | 19.7 ± 4.5 | 16.1 ± 4.1 | 12.3 ± 3.3 | 8.5 ± 2.5 | 6.4 ± 2 |
| Ruoqiang | Train | 9.3 ± 2.5 | 10.9 ± 2.7 | 16 ± 4.3 | 20.1 ± 4.7 | 22.1 ± 5.9 | 22.9 ± 6.6 | 24.1 ± 6.7 | 21.9 ± 6.1 | 19 ± 3.5 | 14.8 ± 3.1 | 9.5 ± 2.7 | 8 ± 1.8 |
| | Test | 8.6 ± 2.6 | 11.2 ± 2.8 | 15.4 ± 3.8 | 18.8 ± 5.2 | 21.8 ± 6 | 23 ± 5 | 21.5 ± 5.9 | 20.3 ± 5.5 | 17.9 ± 4 | 14.2 ± 2.8 | 10.6 ± 2.2 | 7.8 ± 1.9 |
| Khotan | Train | 10.1 ± 2.5 | 11.6 ± 3.3 | 15.5 ± 4 | 19.8 ± 5.4 | 23.4 ± 5.8 | 23.9 ± 6 | 22.3 ± 6.3 | 20.1 ± 5.4 | 18.6 ± 4.8 | 16.3 ± 2.8 | 11.1 ± 2.3 | 8.8 ± 2.6 |
| | Test | 9.1 ± 3 | 11.2 ± 3.8 | 15.2 ± 4.6 | 18.9 ± 5.2 | 21.5 ± 5.1 | 22.1 ± 5.4 | 21.3 ± 5.9 | 19.2 ± 4.6 | 16.2 ± 4.9 | 14.9 ± 2.8 | 10.8 ± 2.2 | 8.7 ± 1.7 |

Note: the unit of the data is MJ $m^{-2}$ $d^{-1}$.

The data were also divided into two parts, from 2006 to 2010 for the training model, and from 2011 to 2015 for validation.

### 2.2. Quantile Mapping (QM)

QM algorithms are commonly used to correct forecasting and observed data [34,35]. The QM method assumes that forecast data has the same cumulative frequency distribution (CDF) as observed data. The general equation of the QM method is defined as follows:

$$\hat{x}_{m,f}(t) = F_{o,h}^{-1}\left\{F_{m,h}\left[x_{m,f}(t)\right]\right\} \tag{1}$$

where $\hat{x}_{m,f}(t)$ is the model forecast data at the $t$ time. $F_{m,h}$ is the CDF of the observed history data. $F_{o,h}^{-1}$ is the inverse of CDF observed historical data.

### 2.3. Equiratio Cumulative Distribution Function Matching (EDCDFm)

EDCDFm is also a method based on quantile mapping. However, unlike the QM method, EDCDFm believes that observed value and forecast value have different CDFs [36]. The difference in CDF function needs to be considered, which is defined as follows:

$$\hat{x}_{m,f}(t) = x_{m,f}(t) + F_{o,h}^{-1}\left\{F_{m,f}\left[x_{m,f}(t)\right]\right\} - F_{m,h}^{-1}\left\{F_{m,f}\left[x_{m,f}(t)\right]\right\} \tag{2}$$

where $F_{m,f}$ is the CDF of the model forecasting data in the future, and $F_{m,h}^{-1}$ is the inverse of CDF model forecasting historical data.

### 2.4. Machine-Learning Algorithms

#### 2.4.1. Long-Short Term Memory (LSTM)

In recent years, due to the advantages of LSTM model in dealing with sequential tasks, researchers have carried out a lot of research on it [37–39]. LSTM is a deep-learning architecture that aims to solve the long-term dependence problem of existing recurrent neural networks (RNNs) by introducing forgetting gates. LSTM model can recall previous data and evaluate the correlation of features based on past data.

As shown in Figure 2, a typical LSTM network consists of one unit and three gates (input gate, forget gate and output gate). The input gate adjusts the amount of new data stored in the unit. The output gate determines which information to obtain from the cell, while the forgetting gate determines which information can be discarded. The LSTM model will consider all this information and make judgments. These gates control cell state Ct and output ht; the input gate can be calculated as follows:

$$gate(f_i) = \sigma_s(w_i x_t + u_i h_{t-1} + b_i) \tag{3}$$

where $\sigma_s$ is the sigmoid activation function, $h_{t-1}$ is the cell output at the previous time step, $w_i$ and $u_i$ are the weighting factors, and $b_i$ is the bias. The forget gate can be calculating as follows:

$$gate(f_t) = \sigma_s(w_f x_t + u_f h_{t-1} + b_f) \tag{4}$$

where $w_f$ and $u_f$ are the weighting factors, and $b_f$ is the bias. The equation of output gate is as follows:

$$gate(f_o) = \sigma_s(w_o x_t + u_o h_{t-1} + b_o) \tag{5}$$

where $w_o$ and $u_o$ are the weighting factors, and $b_f$ is the bias.

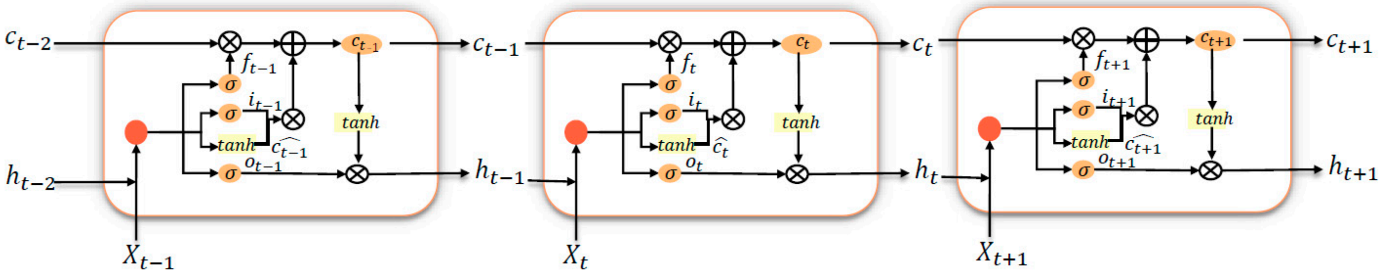

**Figure 2.** The struct of the LSTM model.

In this study, LSTM is used to forecast Rs, and the input data includes $Rs_f$, $Tmax_f$, $Tmin_f$, $RH_f$ and $U_f$ for the forecast target day and observed Rs values during the previous 3–6 days. To achieve this model, Python 3.7 (https://www.python.org/downloads/release/python-370/ (accessed on 25 April 2022)) was used to develop the model.

2.4.2. Support Vector Machine (SVM)

SVM is an advanced statistical method based on the structural risk minimization principle and Vapnik–Chervonenkis dimension theory [40]. This method can be used to deal with classification and regression problems. Support vector regression (SVR) is an extension of a support vector machine in the field of regression. It has the advantages of solid generalization ability and fast convergence speed. It also has the advantages of dealing with small samples and nonlinear problems. By introducing the structural error minimization criterion, SVR has good robustness, generalization and learning ability. The SVR function is defined as follows:

$$f(x) = w\psi(x) + b \tag{6}$$

where $f(x)$ is the output, $w$ is the weight vector, $\psi(x)$ is the high-dimensional nonlinear mapping function, and $b$ is the constant. This equation is equivalent to the following objective function:

$$\min R(F) = \frac{1}{2}\| w \|^2 + C\sum_{i=1}^{n} |f(x_i) - y_i|_\varepsilon \tag{7}$$

where $C$ is the penalty parameter, $n$ is the number of the samples for develop the model, $\varepsilon$ is the maximum allowable error which depending on the samples, and $|f(x_i) - y_i|$ is the residual error, defined as follows:

$$|f(x) - y|_\varepsilon = \max\{0, |f(x) - y| - \varepsilon\} \tag{8}$$

By introducing two relaxation variables ($\zeta$ and $\zeta^*$ ), Equation (5) can be rewritten as:

$$\min \frac{1}{2}\| w \|^2 + C\sum_{i=1}^{n} (\xi_i + \xi_i^*) \tag{9}$$

$$s.t. = \begin{cases} y_i - f(x_i) \le \varepsilon + \xi_i \\ f(x_i) - y_i \le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* > 0 \end{cases} \tag{10}$$

Equation (6) can be converted into a duality problem as:

$$f(x) = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) K(x_i, x_j) + b \tag{11}$$

where $\alpha_i$ and $\alpha_i^*$ are the Lagrange multipliers, and $K(\cdot)$ is a kernel function.

$$K(x_i, x_j) = \exp\left(-\frac{1}{2\sigma^2} \|x_i - x_j\|^2\right) \tag{12}$$

There are many kinds of kernel functions. In this study, we used radial basis function as kernel function, which has certain advantages in nonlinear aspect.

#### 2.4.3. Extreme Gradient Boosting (XGBoost)

XGBoost is the first parallel gradient enhanced tree (GBDT) algorithm. XGBoost, based on classification and regression tree (CART) theory, has been widely proven to be a very efficient approach to regression and classification problems [41]. After optimization, XGBoost's objective function consists of two different parts, representing the deviation and regularization terms of the model to prevent overfitting. The objective function can be written as follows:

$$Obj = \sum_{i=1}^{m} l\left(y_i, \hat{y}_i^{(t-1)} + f_i(x_i)\right) + \Omega(f_k) \tag{13}$$

$$\Omega(f_k) = \gamma T + \frac{1}{2}\lambda \|w\|^2 \tag{14}$$

where $\gamma$ and $\lambda$ are parameters that measure model complexity. $T$ is the number of leaves on the CART tree, and $w$ is also a weight parameter; in XGBoost, this is the weight of each leaf. More details can be found in the reference [41].

#### 2.4.4. Kernel-Based Nonliear Extension of Arps Decline (KNEA)

KNEA is a new time-series model which has been applied in oil-production estimation, $ET_0$ prediction and groundwater-level prediction [42,43]. The KNEA synthesizes nonlinear models of past state and present effects. The main function of KNEA algorithm can be expressed as:

$$f(x) = af(x-1) + g(u(x)) + b \tag{15}$$

where $f(x)$ is the output at present; $f(x-1)$ is the output of the previous step; $u(x)$ is the variables that can affect the output; $g(u(x))$ is a function of variables; and $a$ and $b$ are the constants. Usually, the $g(u(x))$ is unknown, so it should be converted to:

$$g(u(x)) = \omega^T \varphi(u(x)) \tag{16}$$

where $\varphi(u(x))$ is the nonlinear mapping of variable into a new space.

After transformation, a small value $e_x$ can be introduced in the function and the original problem is transformed into a minima problem:

$$e_x = f(x) - af(x-1) - \omega^T \varphi(u(x)) - b \tag{17}$$

$$\min \varsigma(a, \omega, e) = \frac{1}{2}a^2 + \frac{1}{2}\| w \|^2 + \frac{\delta}{2}\sum_{x=2}^{n} e_x^2 \tag{18}$$

$$s.t. f(x) = af(x-1) + g(u(x)) + b + e_x \tag{19}$$

Similar to SVM, this equivalent form can also be solved by introducing Lagrange multipliers and kernel functions.

### 2.4.5. Bat Algorithm

Yang and He [44] proposed the bat algorithm by imitating the predation law of bats. Bat algorithm has high efficiency in parameter optimization. Velocity and position changes are critical for bats to find optimal solutions in space, and these values are obtained by the following equation:

$$f_i = f_{\min} + (f_{\max} - f_{\min})\beta_0 \tag{20}$$

$$V_i^t = V_i^{t+1} + \left(X_i^{t+1} + \left(X_i^t - X^*\right) f_i \tag{21}$$

$$X_i^t = X_i^{t-1} + V_i^t \tag{22}$$

where $\beta_0$ is a random vector and its value range is $[-1, 1]$; $X^*$ is the current optimal position in all the bats; and $f_{\min}$ and $f_{\max}$ are adjust the coefficient of speed. After each generation, each bat produces a new position, as follows:

$$X_{new} = X_{old} + \mu A^t \tag{23}$$

where $\mu$ is also a random vector and its value range is $[-1, 1]$; the following steps implement conditional updates of bat positions. A random number is generated and every bat of this generation is traversed. When the random number is greater than $r_i^t$ and the fitness of the bat is higher than the optimal fitness of the current population, the generated new solution is accepted, and $r_i^t$ as well as $A_i^t$ are updated:

$$A_i^{t+1} = v A_i^{t+1} \tag{24}$$

$$r_i^{t+1} = r_i^0 [1 - e(-\rho t)] \tag{25}$$

where $A_i^1$ = 0.99, and $r_i^0$= $0.5 r_0^i$. Equations (20)–(25) are repeated until the maximum generation is reached. BA has been used to optimize the parameters of the machine-learning models, e.g., SVM, XGBoost and KNEA. In this study, the population of BA algorithm was set to 50 and the number of iterations was 200.

The range of parameters of the three machine-learning models are shown in Table 2.

**Table 2.** Parameters of the three machine-learning models.

| Model | Parameter Names | Range |
|---|---|---|
| SVM | Regularization coefficient | [0.01, 10,000] |
| | Kernel parameter | [0.01, 10,000] |
| XGBoost | Number of trees | [50, 1000] |
| | Maximum tree depth | [2, 50] |
| | Learning rate | [0.01, 0.3] |
| KNEA | Regularization coefficient | [0.1, 10,000] |
| | Kernel parameter | [0.1, 10,000] |

### 2.4.6. Particle Swarm Optimization Algorithm (PSO)

PSO is an algorithm developed by simulating group predation to find the optimal solution [45,46]. The particle swarm optimization algorithm designs a massless particle with only two attributes: speed and position, in which speed represents the speed of movement and position represents the direction of movement. Each particle searches for the optimal solution separately in the search space and records it as the current individual extreme value. The position of the extreme value is shared with other particles in the whole particle swarm. After other particles find the optimal individual extreme value, they update it to the whole particle swarm's current global optimal solution. The formula of position and speed of the PSO algorithm is as follows:

$$Z_i^t = Z_i^{t-1} + u_i^t \tag{26}$$

$$u_i^t = \omega^t \times u_i^{t-1} + c_1 \times \theta_1 \times \left(pbest_i - Z_i^{t-1}\right) + c_2 \times \theta_2 \times \left(gbest_i - Z_i^{t-1}\right) \tag{27}$$

where $Z_i^t$ is the location of the $i$-th particle during $t$-th iteration, $u_i^t$ is the speed of the $i$-th particle during $t$-th iteration, c1 and c2 are study factors and the value was set as 2. $\theta_1$ and $\theta_2$ are random data ranged $[-1, 1]$. $pbest_i$ is the best location of the $i$-th particle among different iterations. $gbest_i$ is the globally best location of all the particles. $\omega^t$ is the momentum factor, which can be calculated as follows:

$$\omega^t = (\omega_{ini} - \omega_{end})(I_{\max} - t)/I_{\max} + \omega_{end} \tag{28}$$

where $\omega_{ini}$ and $\omega_{end}$ are the initial and the end momentum factors, and the values were set as 0.9 and 0.4, respectively. $I_{\max}$ is the maximum iteration. In this study, the population of PSO algorithm was set to 50 and the number of iterations was 200. The range of machine-learning parameters optimized by PSO is the same as the BA algorithm. Figure 3 presents the flow chart of three machine-learning models optimized by evolutionary algorithms. To achieve these models, except LSTM, R language (R language v4.4, https://www.r-project.org/ (accessed on 25 April 2022)) was used.

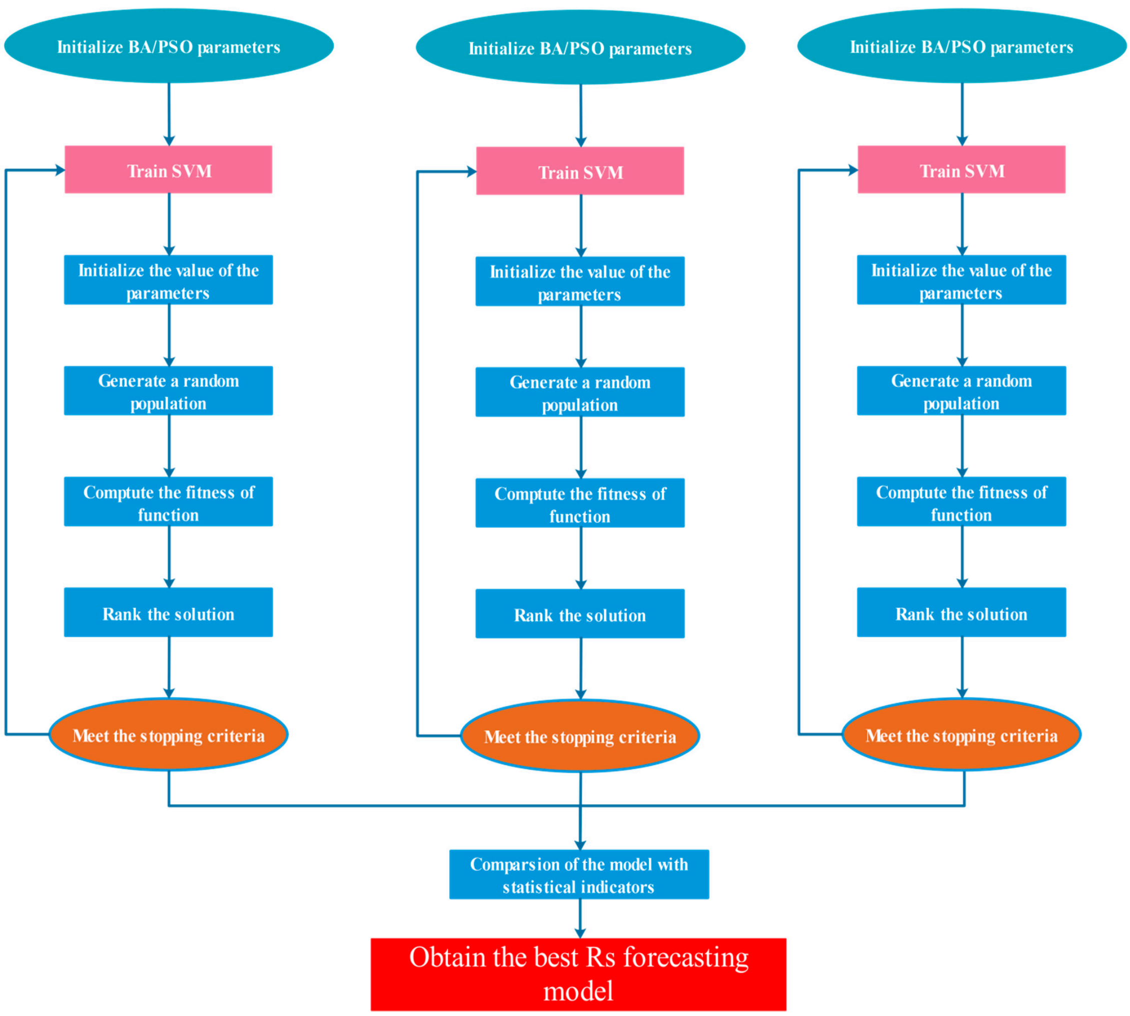


**Figure 3.** Flowchart of the three machine-learning models optimal for evolutionary algorithms.

### 2.5. Statistical Indicators

In this study, four commonly used statistical indicators were used to evaluate the prediction performance of total surface radiation, which are determination coefficient ($R^2$):

$$R^2 = \frac{\left[\sum_{i=1}^{n} (R_{s,m} - \overline{R}_{s,m})\left(R_{s,f} - \overline{R}_{s,f}\right)\right]^2}{\sum_{i=1}^{n} (R_{s,m} - \overline{R}_{s,m})2\sum_{i=1}^{n}\left(R_{s,f} - \overline{R}_{s,f}\right)^2} \quad (29)$$

Root mean square error (RMSE):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(R_{s,m} - R_{s,f}\right)^2} \quad (30)$$

Mean absolute error:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|R_{s,m} - R_{s,f}\right| \quad (31)$$

Normalized *RMSE*:

$$NRMSE = RMSE/\overline{R}_{s,m} \quad (32)$$

where Rs,m is measured Rs, Rs,f is forecasting Rs , $\overline{R}_{s,m}$ is the mean of the measured Rs, and $\overline{R}_{s,f}$ is the mean of forecasting Rs.

## 3. Results

### 3.1. Empirical Statistics Methods

Table 3 presents the statistical indicators of the GEFSv12 NWP raw $Rs_f$ forecasting data and the results from QM and EDCDFm methods. In general, with the extension of the forecast period, the errors of NWP raw $Rs_f$ data and the $Rs_f$ correct by QM and EDCDFm methods gradually increase. In Altay, the performance of the QM and EDCDFm methods were very similar, and both $o_f$ them were slightly better than that of the NWP raw $Rs_f$ data. In Kashgar, the error of the raw $Rs_f$ data was relatively large. However, the QM and EDCDFm methods were superior to the raw $Rs_f$, with RMSE decreased by 28.2–31% and 28.6–31.5%, and MAE decreased by 27.9–31.1% and 27.7–31.1%, respectively, during 1–3 d ahead. In Ruoqiang, the error of the raw $Rs_f$ was large, and its RMSE was more than 5 MJ $m^{-2}$ $d^{-1}$. After correcting by QM and EDCDFm method correction, RMSE decreased by 17.4–18.5% and 19.7–20.1% for 1–3 d ahead, and MAE decreased by 16–17.7% and 17.7–19.4%, respectively. However, the $R^2$ of the raw $Rs_f$ was slightly higher than that of the two statistical methods. This indicates that the statistical method improved the overestimation (or underestimation) problem. The performance for Khotan station was similar to that of Ruoqiang station. Compared with the raw $Rs_f$ over the four stations, the RMSE and MAE of QM and EDCDFm models decreased by 20% and 15%, respectively. It can be seen from the above results that empirical statistical methods can improve forecasting accuracy.

As can be seen from the scatter plot of raw Rs vs. ground observed Rs (Figure 4), the discrete points increased slightly from 1 d to 3 d, indicating a slight decrease in inaccuracy. The forecasting value of $Rs_f$ in the future 1–3 d was not higher than 30 MJ $m^{-2}$ $d^{-1}$, which was slightly lower than the extreme value of Rs. The main problem of the GEFS data set lay in the existence of many overestimated discrete points when the observed value was lower than 25 MJ $m^{-2}$ $d^{-1}$. However, the QM and EDCDFm methods can alleviate this problem, and the $R^2$ of the two methods was slightly higher than the corresponding value of raw $Rs_f$ data.

**Table 3.** Statistical indicators of solar-radiation forecast by GEFS NWP raw data and two empirical-statistics methods.

| ID | 1 d | | | | 2 d | | | | 3 d | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Model** | **$R^2$** | **RMSE** | **MAE** | **NRMSE** | **$R^2$** | **RMSE** | **MAE** | **NRMSE** | **$R^2$** | **RMSE** | **MAE** | **NRMSE** |
| 51076 Altay | | | | | | | | | | | | |
| NWP | 0.816 | 3.939 | 3.120 | 0.250 | 0.766 | 4.313 | 3.417 | 0.274 | 0.745 | 4.582 | 3.482 | 0.292 |
| QM | 0.821 | 3.843 | 3.077 | **0.246** | 0.787 | 4.194 | 3.304 | 0.269 | **0.768** | 4.387 | 3.442 | **0.281** |
| EDCDFm | **0.820** | **3.838** | **3.071** | **0.246** | **0.788** | **4.189** | **3.301** | **0.268** | **0.768** | **4.384** | **3.437** | **0.281** |
| 51709 Kashgar | | | | | | | | | | | | |
| NWP | 0.795 | 5.016 | 3.822 | 0.327 | 0.772 | 5.214 | 3.955 | 0.340 | 0.757 | 5.378 | 4.080 | 0.351 |
| QM | 0.816 | 3.460 | **2.633** | 0.217 | 0.792 | 3.707 | **2.798** | 0.233 | 0.776 | 3.862 | **2.943** | 0.243 |
| EDCDFm | **0.820** | **3.437** | **2.633** | **0.216** | **0.795** | **3.699** | 2.815 | **0.232** | **0.780** | **3.841** | 2.950 | **0.241** |
| 51777 Ruoqiang | | | | | | | | | | | | |
| NWP | 0.753 | 4.547 | 3.102 | 0.280 | **0.726** | 4.859 | 3.312 | 0.299 | **0.697** | 5.156 | 3.478 | 0.317 |
| QM | 0.754 | 3.708 | 2.553 | 0.224 | 0.713 | 4.002 | 2.762 | 0.241 | 0.681 | 4.257 | 2.920 | 0.257 |
| EDCDFm | **0.758** | **3.632** | **2.499** | **0.219** | 0.719 | **3.912** | **2.709** | **0.236** | 0.688 | **4.138** | **2.864** | **0.250** |
| 51828 Khotan | | | | | | | | | | | | |
| NWP | 0.701 | 4.822 | 3.398 | 0.296 | 0.668 | 5.127 | 3.628 | 0.315 | 0.650 | 5.354 | 3.788 | 0.329 |
| QM | 0.720 | 3.674 | 2.750 | 0.219 | 0.665 | 4.057 | 3.048 | 0.241 | 0.649 | 4.178 | **3.143** | 0.249 |
| EDCDFm | **0.721** | **3.637** | **2.733** | **0.216** | **0.669** | **4.012** | **3.044** | **0.239** | **0.652** | **4.145** | 3.151 | **0.247** |

Note: the value in bold is the best statistical indicator among the different methods. The same as follow.

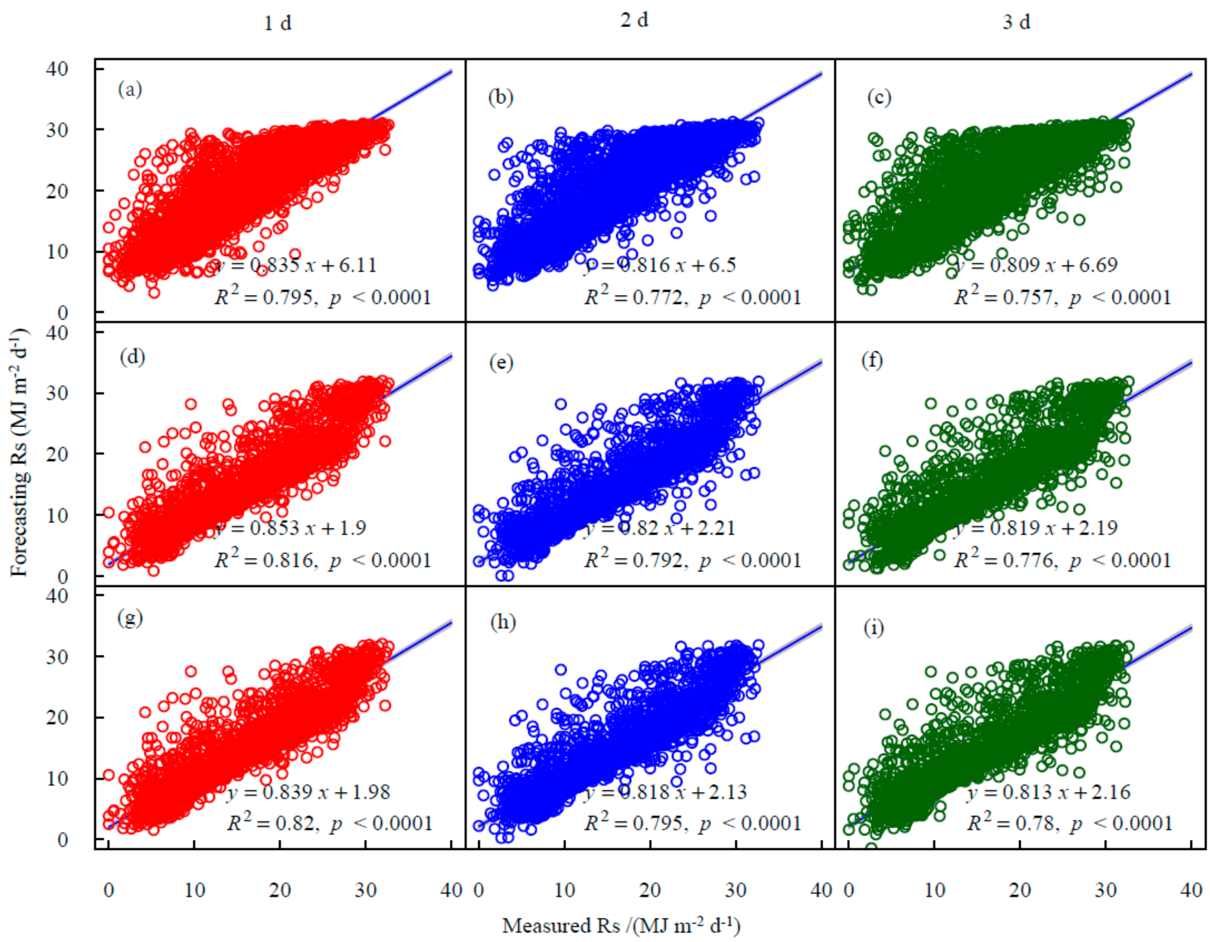


**Figure 4.** Scatter plots of measured Rs vs. forecasting Rs at Kashgar station during the testing period, GEFSv12 raw Rs forecasting data on (**a**) 1 d ahead, (**b**) 2 d ahead, (**c**) 3 d ahead; QM method forecasting Rs on (**d**) 1 d ahead, (**e**) 2 d ahead, (**f**) 3 d ahead; EDCDFm method forecasting Rs on (**g**) 1 d ahead, (**h**) 2 d ahead, (**i**) 3 d ahead.

### 3.2. Machine-Learning Methods

Table 4 shows the statistical indicators of Rs-forecasting results by seven different machine-learning methods during 1–3 d ahead. In Altay, the third day's $R^2$ average increased by 0.046, and the average RMSE and MAE increased by 13.4% and 13.1%, compared with the first day. Among the seven machine-learning models, the BA-KNEA model was superior to other machine−learning models each day, and the RMSE, MAE and NRMSE of the BA-KNEA model decreased by 2.1–10.3%, 2.5–12.0% and 2.8–12.4% than other machine-learning models for 1 d ahead, decreased by 1.8–8.8%, 1.7% to 10.1% and 1.6–9.9% for 2 d ahead, and decreased by 2.2–8.2%, 2.2–9.6% and 2.0–9.5% for 3 d ahead. The performance of the BA-SVM model was ranked second, followed by BA-XGBoost, PSO-KNEA, PSO-SVM, LSTM and PSO-XGBoost models.

**Table 4.** Statistical indicators of solar-radiation forecasts by different machine-learning models.

| ID | 1 d | | | | 2 d | | | | 3 d | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Model** | $R^2$ | **RMSE** | **MAE** | **NRMSE** | $R^2$ | **RMSE** | **MAE** | **NRMSE** | $R^2$ | **RMSE** | **MAE** | **NRMSE** |
| 51076 Altay | | | | | | | | | | | | |
| LSTM | 0.813 | 3.889 | 3.086 | 0.202 | 0.798 | 4.178 | 3.314 | 0.216 | 0.787 | 4.258 | 3.168 | 0.207 |
| PSO-SVM | 0.817 | 3.875 | 2.988 | 0.191 | 0.792 | 4.116 | 3.181 | 0.204 | 0.773 | 4.292 | 3.319 | 0.213 |
| BA-SVM | 0.837 | 3.627 | 2.854 | 0.183 | 0.811 | 3.91 | 3.032 | 0.194 | 0.793 | 4.091 | 3.174 | 0.203 |
| PSO-XGBoost | 0.816 | 3.917 | 3.118 | 0.2 | 0.79 | 4.178 | 3.28 | 0.21 | 0.773 | 4.33 | 3.403 | 0.218 |
| BA-XGBoost | 0.833 | 3.685 | 2.893 | 0.185 | 0.803 | 4.005 | 3.114 | 0.199 | 0.786 | 4.171 | 3.243 | 0.208 |
| PSO-KNEA | 0.826 | 3.723 | 2.903 | 0.186 | 0.794 | 4.053 | 3.088 | 0.198 | 0.77 | 4.281 | 3.268 | 0.209 |
| BA-KNEA | **0.844** | **3.552** | **2.785** | **0.178** | **0.819** | **3.839** | **2.98** | **0.191** | **0.803** | **4.002** | **3.105** | **0.199** |
| 51709 Kashgar | | | | | | | | | | | | |
| LSTM | 0.834 | 3.485 | 2.81 | 0.177 | 0.808 | 3.735 | 2.75 | 0.173 | 0.799 | 3.908 | 3.033 | 0.191 |
| PSO-SVM | 0.838 | 3.436 | 2.641 | 0.166 | 0.809 | 3.735 | 2.863 | 0.18 | 0.789 | 3.824 | 2.886 | 0.181 |
| BA-SVM | 0.861 | 3.38 | 2.707 | 0.17 | 0.838 | 3.596 | 2.854 | 0.179 | 0.799 | 3.923 | 3.136 | 0.197 |
| PSO-XGBoost | 0.84 | 3.445 | 2.7 | 0.17 | 0.811 | 3.754 | 2.933 | 0.184 | 0.8 | 3.808 | 2.982 | 0.187 |
| BA-XGBoost | 0.845 | 3.345 | 2.55 | 0.16 | 0.819 | 3.661 | 2.775 | 0.174 | 0.808 | 3.677 | 2.796 | 0.176 |
| PSO-KNEA | 0.841 | 3.231 | 2.438 | 0.153 | 0.824 | 3.45 | **2.629** | **0.165** | 0.801 | 3.618 | 2.748 | 0.173 |
| BA-KNEA | **0.869** | **3.056** | **2.37** | **0.149** | **0.837** | **3.434** | 2.654 | 0.167 | **0.834** | **3.487** | **2.733** | **0.172** |
| 51777 Ruoqiang | | | | | | | | | | | | |
| LSTM | 0.784 | 3.401 | 2.431 | 0.147 | 0.74 | 3.821 | 2.547 | 0.159 | 0.719 | 3.852 | 2.749 | 0.168 |
| PSO-SVM | 0.796 | 3.313 | 2.331 | 0.141 | 0.76 | 3.603 | 2.528 | 0.153 | 0.732 | 3.796 | 2.711 | 0.164 |
| BA-SVM | 0.803 | 3.266 | 2.296 | 0.139 | 0.764 | 3.592 | 2.542 | 0.153 | 0.733 | 3.811 | 2.693 | 0.163 |
| PSO-XGBoost | 0.787 | 3.423 | 2.429 | 0.147 | 0.75 | 3.688 | 2.614 | 0.158 | 0.731 | 3.822 | 2.746 | 0.166 |
| BA-XGBoost | 0.796 | 3.319 | 2.304 | 0.139 | 0.753 | 3.639 | 2.542 | 0.153 | 0.721 | 3.853 | 2.728 | 0.165 |
| PSO-KNEA | 0.785 | 3.552 | 2.41 | 0.145 | 0.739 | 3.86 | 2.6 | 0.157 | 0.717 | 4.069 | 2.736 | 0.165 |
| BA-KNEA | **0.819** | **3.123** | **2.196** | **0.133** | **0.791** | **3.354** | **2.387** | **0.144** | **0.752** | **3.674** | **2.624** | **0.158** |
| 51828 Khotan | | | | | | | | | | | | |
| LSTM | 0.762 | 3.331 | 2.619 | 0.155 | 0.717 | 3.739 | 2.74 | 0.161 | 0.696 | 3.873 | 2.822 | 0.166 |
| PSO-SVM | 0.752 | 3.459 | 2.665 | 0.159 | 0.71 | 3.731 | 2.815 | 0.167 | 0.697 | 3.81 | 2.883 | 0.172 |
| BA-SVM | 0.771 | 3.384 | 2.664 | 0.159 | 0.737 | 3.755 | 2.968 | 0.177 | 0.704 | 3.969 | 3.116 | 0.185 |
| PSO-XGBoost | 0.755 | 3.32 | 2.621 | 0.151 | 0.723 | 3.885 | 3.003 | 0.179 | 0.703 | 3.991 | 3.197 | 0.189 |
| BA-XGBoost | 0.754 | 3.37 | 2.678 | 0.157 | 0.734 | 3.689 | 2.847 | 0.167 | 0.722 | 3.788 | 2.895 | 0.175 |
| PSO-KNEA | 0.743 | 3.506 | 2.587 | 0.154 | 0.689 | 3.834 | 2.8 | 0.167 | 0.671 | 3.929 | 2.899 | 0.172 |
| BA-KNEA | **0.783** | **3.227** | **2.509** | **0.149** | **0.754** | **3.483** | **2.676** | **0.159** | **0.737** | **3.576** | **2.732** | **0.163** |

In Kashgar, the BA-KNEA model did not have a significant advantage over the PSO-KNEA model on the first two days, but performed slightly better than the PSO-KNEA model on the third day. In addition, the BA-KNEA model was generally superior to other models. RMSE, MAE and NRMSE decreased by 3.8–7.0%, 0–6.2% and 0–6.3% for 1 d ahead, 3.8–8.4%, 2.8–8.6% and 2.4–8.2% for 2 d ahead, and 5.4–12.5%, 2.3–14.7% and 2.3–14.5% for 3 d ahead. In addition, the BA-XGBoost model slightly outperformed the BA-SVM model.

In Ruoqiang, the BA-KNEA model performed better than the other six models. Compared with the BA-KNEA model, the RMSE, MAE and NRMSE of the other six models increased by 4.6–9.6%, 4.6–10.6% and 4.5–10.5% for 1 d ahead, 7.1–10.0%, 4.9–9.5%, 6.3–9.7%

for 2 d ahead, and 3.3–4.9%, 2.6–4.6%, 3.2–5.1% for 3 d ahead. The BA-SVM model performed better than the other four models on 1 d, but the advantage in the other models, except BA-KNEA, was not obvious on the other two days. In Khotan, the BA-KNEA model also achieved the highest accuracy, and the RMSE, MAE and NRMSE of the other five models increased by 2.6–6.9%, 4.8–7.1% and 1.3–6.7% for 1 d ahead, 3.5–9.0%, 1.6–8.4%, 1.2–8.5% for 2 d ahead, and 3.0–8.5%, 1.8–12.8%, 1.8–11.8% for 3 d ahead. The performance of the BA-SVM model was still better than the other four models, except for the BA-KNEA model.

The scatter plots of observed Rs vs. $Rs_f$ by seven machine-learning models are shown in Figure 5. Among all the machine-learning models, it can be seen that the BA-KENA model performed slightly better than other models, followed by the BA-SVM model. The slope of all the regression equations in the Figure was less than 1, and the intercept was greater than 0, which means that all the models exhibit the problem that when Rs is very large, the model will underestimate the result, and when Rs is very small, the model will underestimate the result.

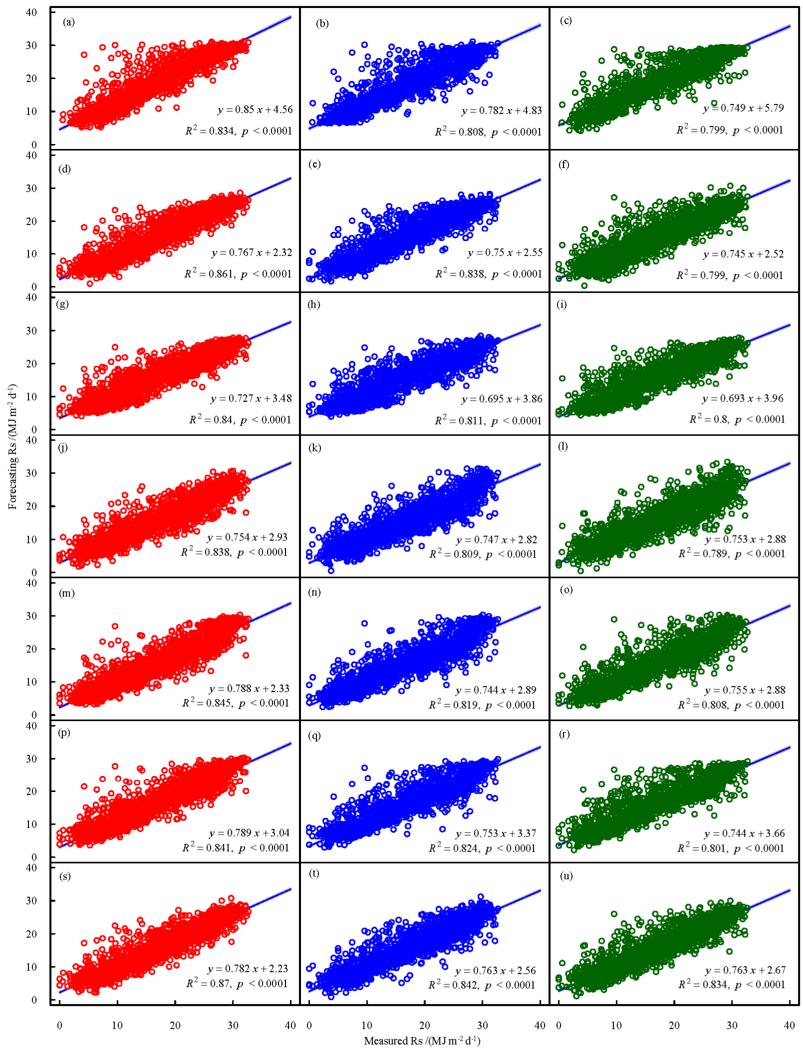


**Figure 5.** Scatter plots of measured Rs vs. forecasting Rs at Kashgar station during the testing period, LSTM method forecasting data on (**a**) 1 d ahead, (**b**) 2 d ahead, (**c**) 3 d ahead; PSO-SVM method forecasting Rs on (**d**) 1 d ahead, (**e**) 2 d ahead, (**f**) 3 d ahead; BA-SVM method forecasting Rs on (**g**) 1 d ahead, (**h**) 2 d ahead, (**i**) 3 d ahead; PSO-XGBoost method forecasting Rs on (**j**) 1 d ahead, (**k**) 2 d ahead, (**l**) 3 d ahead; BA-XGBoost method forecasting Rs on (**m**) 1 d ahead, (**n**) 2 d ahead, (**o**) 3 d ahead; PSO-KNEA method forecasting Rs on (**p**) 1 d ahead, (**q**) 2 d ahead, (**r**) 3 d ahead; BA-KNEA method forecasting Rs on (**s**) 1 d ahead, (**t**) 2 d ahead, (**u**) 3 d ahead.

Figure 6 shows the distribution of the absolute error of the forecast Rs for different machine-learning models 1–3 days ahead. As can be seen, at 1 d ahead, the proportion of days with AE < 2 MJ $m^{-2}$ $d^{-1}$ for the six models was around 60%; the proportion of PSO-KNEA and BA-KNEA was slightly higher than in other models; and had a AE > 6 MJ $m^{-2}$ $d^{-1}$ days ratio, the BA-KNEA had a slight advantage over the other models. The performance on 2 d ahead was slightly worse than that on 1 d ahead: the proportion of days with AE < 2 MJ $m^{-2}$ $d^{-1}$ for all six models was below 60%, while the number of days with AE > 6 MJ $m^{-2}$ $d^{-1}$ showed little change compared with 1 d ahead, with the BA-KNEA model having a slight advantage over the other models in the number of days with AE > 6 MJ $m^{-2}$ $d^{-1}$. In the 3 d ahead, the accuracy of the six models continued to decline compared with the previous 2 d, and the BA-KNEA model had a slightly lower proportion of days with AE > 6 MJ $m^{-2}$ $d^{-1}$ than the other models.

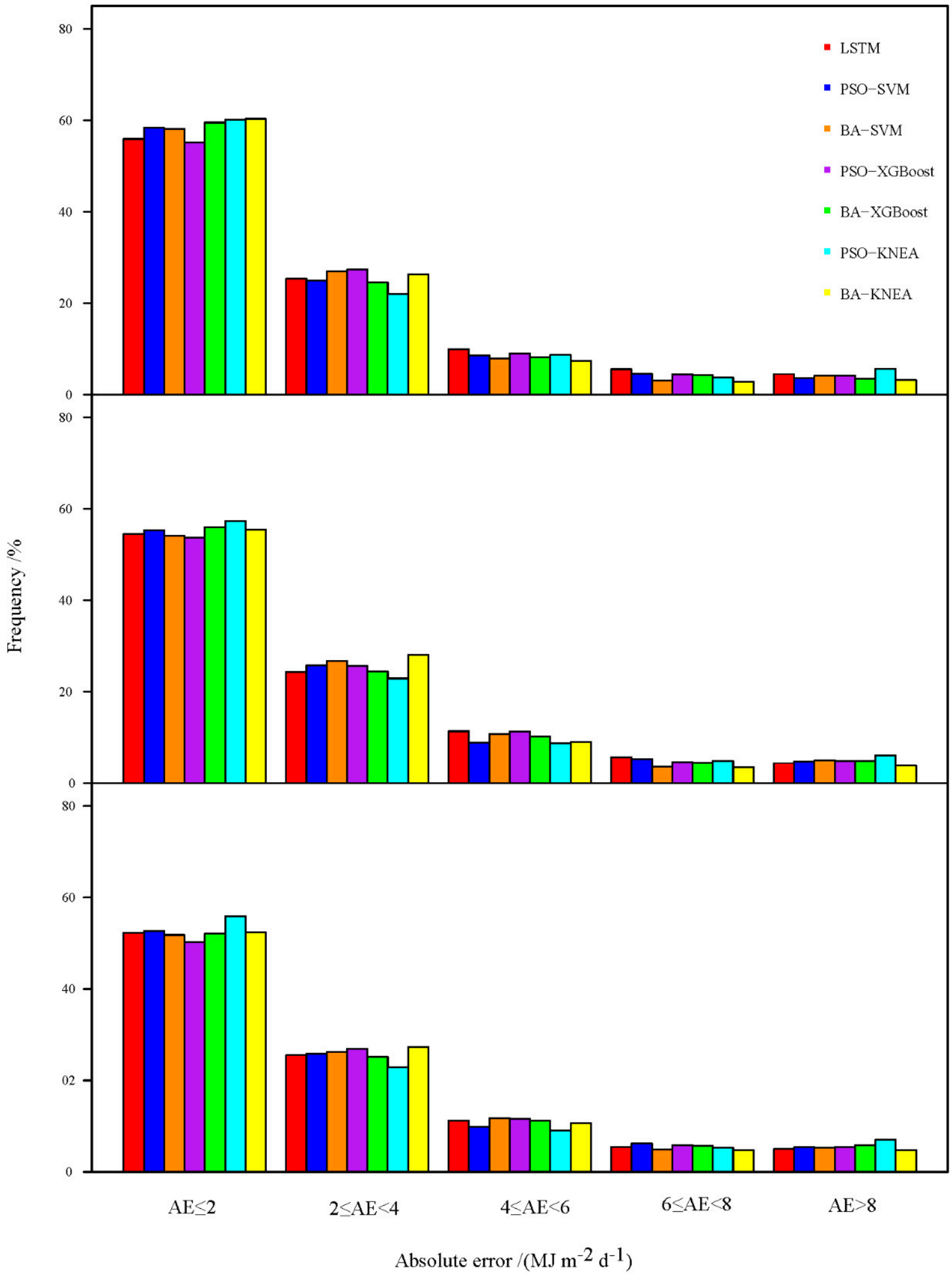


**Figure 6.** Absolute error of different machine−learning models at Ruoqiang station.

Figure 7 shows the Taylor diagram of different methods over the four stations. It can be seen that the BA-KNEA model outperformed the other methods over the all stations.

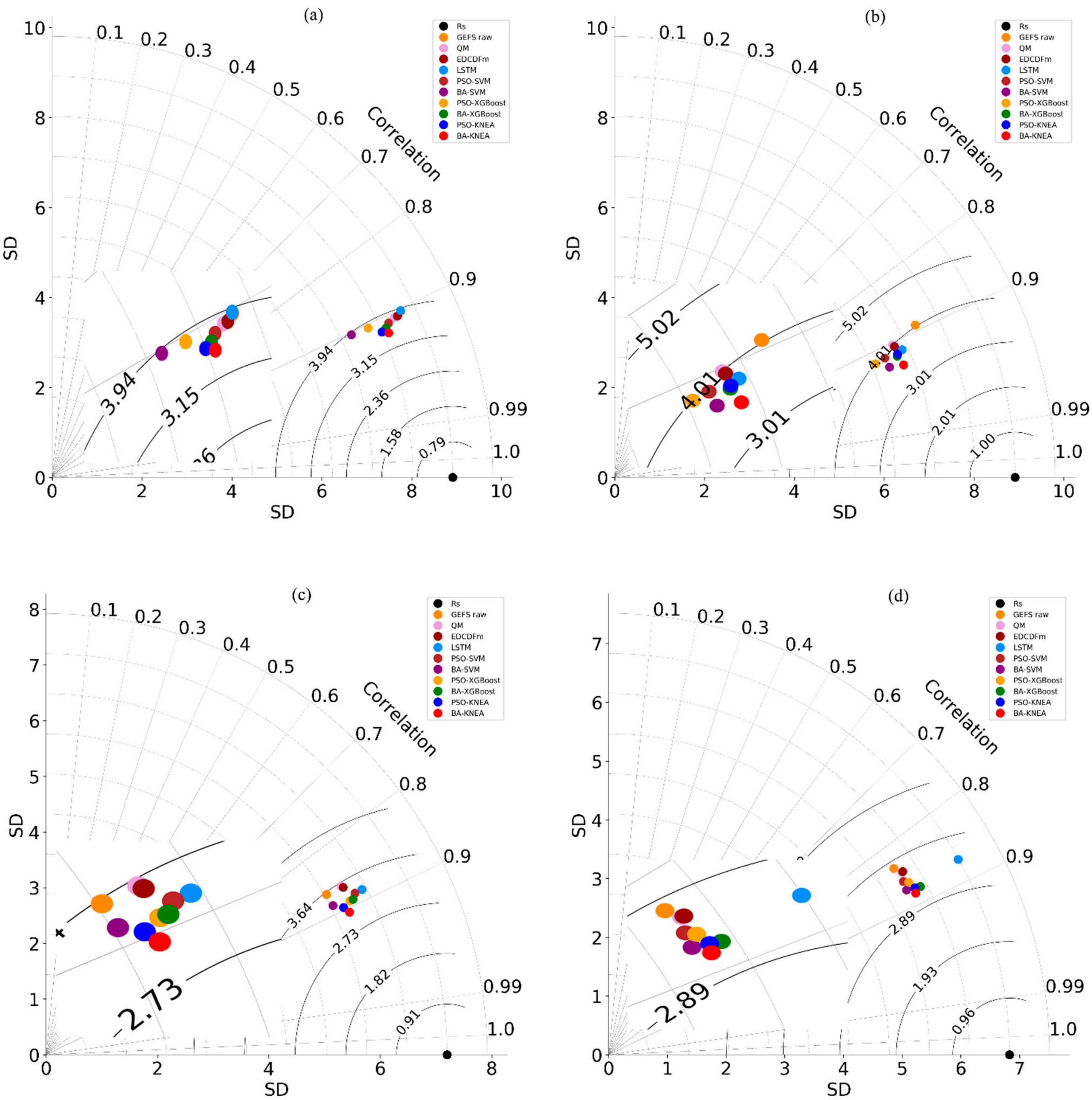


**Figure 7.** Taylor plots of forecasting results for different Rs.

### *3.3. Comparison of Statistical Models and Machine-learning Models*

To evaluate the performance of different categories of models, we ranked the four statistical indicators of all models over the four stations (Table 5). With the highest $R^2$ or the lowest RMSE, MAE or NRMSE would rank first, and so on. When the ranking of different statistical indicators is different, the model with more indicators at the top ranks first. It can be seen that the rank of different models in 1–3 d ahead were the same. The BA-KNEA model was the best, followed by the BA-SVM, BA-XGBoost, PSO-KNEA, PSO-SVM, LSTM, PSO-XGBoost, EDCDFm and QM models. The above results prove that the machine−learning model is superior to the empirical-statistical model, and the new BA-KNEA model has the best performance in accuracy. In addition, the Taylor plots of different stations on the first day of the forecast period are shown in Figure 6. It can also be seen that the results of the BA-KNEA model were the closest to the observations, while the GEFS raw data had the largest error.

**Table 5.** Rank of empirical-statistical and machine-learning models.

| Model | 1 d | 2 d | 3 d |
|---|---|---|---|
| GEFS raw | 10 | 10 | 10 |
| QM | 9 | 9 | 9 |
| EDCDFm | 8 | 8 | 8 |
| LSTM | 6 | 6 | 6 |
| PSO-SVM | 5 | 5 | 5 |
| BA-SVM | 2 | 2 | 3 |
| PSO-XGBoost | 7 | 7 | 7 |
| BA-XGBoost | 3 | 3 | 2 |
| PSO-KNEA | 4 | 4 | 4 |
| BA-KNEA | 1 | 1 | 1 |

### *3.4. BA-KNEA with Different Input Combinations*

In order to analyze the difference in the forecasting ability of different meteorological factors on the results, we used the BA-KNEA model to set up different input combinations. Through the results, we explored the contribution differences of different factors. Table 6 shows the statistical indicators of the different input combinations of the BA-KNEA model in the forecast period 1–3 d. When the input factor is $Rs_f$, the accuracy of the BA-KNEA model was better than that of the QM and EDCDFm methods with the same input at four stations (Table 3), and the RMSE and MAE of the BA-KNEA model was 1.7–7.9% and was 1.6–7.6% lower in the forecast period of 1–3 days, relative to the EDCDFm method. This model was also better than the model established with temperature and extraterrestrial radiation as inputs (Combination 5), which shows that the solar radiation accuracy of the GEFSv12 dataset is better than that of the traditional temperature-based machine-learning model method. In Altay, when only the maximum and minimum air temperature was used as input, the error was larger than the model with Rs input: $R^2$ was between 0.712–0.723, RMSE was between 4.705–4.812 MJ $m^{-2}$ $d^{-1}$, and MAE was between 3.766–3.799 MJ $m^{-2}$ $d^{-1}$, and NRMSE was between 0.241–0.243. Adding $RH_f$, $U_f$, $Tmax_f$ and $Tmin_f$ based on the $Rs_f$ can improve the prediction accuracy of Rs, among which the increase in wind speed was the largest, followed by air temperature, and, finally, relative humidity. Compared with Combination 2, 3, and 4, the accuracy of combination 6 was higher, and it can be seen that the accuracy of the multi-factor was higher than that of the two-factor combination. This shows that the multi-factor combination contains more nonlinear information related to Rs than the two-factor combination, which helps improve the model accuracy further. At Kashgar station, adding relative humidity based on Rs did not improve the accuracy significantly, and when the forecast period was 2 and 3 days, adding wind speed based on Rs slightly improved the accuracy. Adding the temperature model based on Rs improves the model's accuracy to a certain extent, but it is not much different from the accuracy of the complete combination (Combination 6). This is mainly due to the limited contribution of RH and U to improving the accuracy of the model. The performance of the BA-KNEA model on the first two days of Ruoqiang Station was similar to that on Altay, but on the third day, Combination 3 outperformed the complete input combination. Due to poor forecast accuracy of wind speed and relative humidity, adding these factors will increase the noise in the model. At Khotan station, on the first day, the complete combination was close to the Combination 2, 3, and 4 but superior to those during the other two days. The complete combination is slightly better than the other combinations. As seen from the above, the complete combination was slightly better than the other combinations over the four stations.

**Table 6.** Statistical indicators of BA-KNEA model under different input combinations.

| ID | Input | 1 d | | | | 2 d | | | | 3 d | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $R^2$ | RMSE | MAE | NRMSE | $R^2$ | RMSE | MAE | NRMSE | $R^2$ | RMSE | MAE | NRMSE |
| | 51076 Altay | | | | | | | | | | | | |
| 1 | $Rs_f$ | 0.824 | 3.778 | 3.019 | 0.193 | 0.789 | 4.139 | 3.23 | 0.207 | 0.771 | 4.312 | 3.382 | 0.217 |
| 2 | $Rs_f$, $RH_f$ | 0.828 | 3.741 | 2.978 | 0.191 | 0.799 | 4.051 | 3.164 | 0.203 | 0.781 | 4.226 | 3.305 | 0.212 |
| 3 | $Rs_f$, $Tmax_f$, $Tmin_f$ | 0.832 | 3.687 | 2.913 | 0.187 | 0.805 | 3.983 | 3.079 | 0.197 | 0.787 | 4.178 | 3.269 | 0.209 |
| 4 | $Rs_f$, $U_f$ | 0.835 | 3.651 | 2.862 | 0.183 | 0.808 | 3.943 | 3.057 | 0.196 | 0.792 | 4.097 | 3.169 | 0.203 |
| 5 | $Tmax_f$, $Tmin_f$, Ra | 0.723 | 4.705 | 3.766 | 0.241 | 0.721 | 4.757 | 3.737 | 0.239 | 0.712 | 4.812 | 3.799 | 0.243 |
| 6 | All | **0.844** | **3.552** | **2.785** | **0.178** | **0.819** | **3.839** | **2.98** | **0.191** | **0.803** | **4.002** | **3.105** | **0.199** |
| | 51709 Kashgar | | | | | | | | | | | | |
| 1 | $Rs_f$ | 0.852 | 3.21 | 2.499 | 0.157 | 0.829 | 3.494 | 2.711 | 0.17 | 0.814 | 3.663 | 2.87 | 0.18 |
| 2 | $Rs_f$, $RH_f$ | 0.859 | 3.23 | 2.551 | 0.16 | 0.84 | 3.456 | 2.705 | 0.17 | 0.823 | 3.632 | 2.869 | 0.18 |
| 3 | $Rs_f$, $Tmax_f$, $Tmin_f$ | 0.867 | 3.185 | 2.535 | 0.159 | **0.846** | **3.388** | **2.634** | **0.165** | 0.832 | 3.488 | 2.741 | 0.172 |
| 4 | $Rs_f$, $U_f$ | **0.87** | 3.223 | 2.55 | 0.16 | 0.841 | 3.464 | 2.701 | 0.17 | 0.826 | 3.502 | **2.705** | **0.17** |
| 5 | $Tmax_f$, $Tmin_f$, Ra | 0.796 | 3.958 | 3.09 | 0.194 | 0.785 | 3.809 | 2.93 | 0.184 | 0.776 | 3.838 | 2.954 | 0.186 |
| 6 | All | 0.869 | **3.056** | **2.37** | **0.149** | 0.837 | 3.434 | 2.654 | 0.167 | **0.834** | **3.487** | 2.733 | 0.172 |
| | 51777 Ruoqiang | | | | | | | | | | | | |
| 1 | $Rs_f$ | 0.789 | 3.403 | 2.352 | 0.142 | 0.755 | 3.64 | 2.504 | 0.151 | 0.73 | 3.818 | 2.663 | 0.161 |
| 2 | $Rs_f$, $RH_f$ | 0.798 | 3.302 | 2.286 | 0.138 | 0.767 | 3.527 | 2.479 | 0.15 | 0.741 | 3.732 | 2.639 | 0.159 |
| 3 | $Rs_f$, $Tmax_f$, $Tmin_f$ | 0.811 | 3.199 | 2.296 | 0.139 | 0.782 | 3.467 | 2.467 | 0.149 | **0.756** | **3.649** | **2.616** | **0.158** |
| 4 | $Rs_f$, $U_f$ | 0.814 | 3.222 | 2.245 | 0.135 | 0.774 | 3.511 | 2.445 | 0.148 | 0.74 | 3.746 | 2.649 | 0.16 |
| 5 | $Tmax_f$, $Tmin_f$, Ra | 0.745 | 3.764 | 2.792 | 0.168 | 0.724 | 3.875 | 2.871 | 0.173 | 0.702 | 4.035 | 2.96 | 0.179 |
| 6 | All | **0.819** | **3.123** | **2.196** | **0.133** | **0.791** | **3.354** | **2.387** | **0.144** | 0.752 | 3.674 | 2.624 | **0.158** |
| | 51828 Khotan | | | | | | | | | | | | |
| 1 | $Rs_f$ | 0.747 | 3.44 | 2.607 | 0.155 | 0.694 | 3.787 | 2.827 | 0.168 | 0.671 | 3.948 | 2.998 | 0.178 |
| 2 | $Rs_f$, $RH_f$ | 0.769 | 3.293 | 2.523 | 0.15 | 0.719 | 3.645 | 2.767 | 0.165 | 0.705 | 3.729 | 2.818 | 0.168 |
| 3 | $Rs_f$, $Tmax_f$, $Tmin_f$ | 0.782 | 3.236 | **2.5** | **0.149** | 0.751 | 3.564 | 2.771 | 0.165 | 0.731 | 3.678 | 2.833 | 0.169 |
| 4 | $Rs_f$, $U_f$ | 0.763 | 3.337 | 2.504 | **0.149** | 0.725 | 3.643 | 2.786 | 0.166 | 0.708 | 3.765 | 2.867 | 0.171 |
| 5 | $Tmax_f$, $Tmin_f$, Ra | 0.73 | 3.602 | 2.775 | 0.165 | 0.716 | 3.718 | 2.857 | 0.17 | 0.697 | 3.823 | 2.919 | 0.174 |
| 6 | All | **0.783** | **3.227** | 2.509 | **0.149** | **0.754** | **3.483** | **2.676** | **0.159** | **0.737** | **3.576** | **2.732** | **0.163** |

## 4. Discussion

Different machine-learning models perform differently in solar-radiation prediction. This is mainly due to two reasons. Firstly, different machine-learning models have different sensitivities to data distribution. For example, kernel-based machine-learning methods can perform well in low-dimensional data sets [47]. However, the tree-based model performs better with high dimensions and a large amount of typed data. The deep-learning model has better performance in image processing [48]. Another reason is that the parameter selection of machine-learning models did not achieve the optimal global solution. Fan et al. [31] compared the performance of SVM and XGBoost when the input factors were temperature and precipitation and found that SVM was slightly better than the XGBoost model. Ghimire et al. [7] compared ANN, SVR, GPML and GP models for forecasting solar radiation with reanalysis data in Queensland, Australia. They highlighted that the ANN model outperformed other ML models. Shin et al. [49] used a deep-learning model to short-term forecast solar radiation for photovoltaic power generation. Hu et al. [50] used ground-based images and an ANN model to forecast solar radiation. However, there is limited study of using weather-forecast products to forecast solar radiation in China. In this study, we evaluated the capability of the GEFSv12 product in the solar-resource-rich region of China. We found that the raw solar-radiation forecast data in GEFSv12 has poor performance and uncertainty for indirect use. Thus, we built a coupling model based on the bat algorithm and KNEA model. The result shows that the newly developed model is superior to other empirical-statistical and machine-learning models. The LSTM had been used to forecast Rs on hourly and other time scales [51,52]. However, we found that the LSTM did not perform better than the BA-KNEA model nor other models. The daily Rs fluctuated widely on an hourly scale in the arid regions of the northwest of China, and historical information is not as important as the WRF data for future. Thus, the LSTM did not achieve enough information to forecast 1–3 d Rs.

Many scholars have found that various meteorological factors, such as air temperature, relative humidity, wind speed, and precipitation, are closely related to solar radiation [53,54], but the effects of these factors vary in different regions of the globe [55,56]. In northwest China, air temperature is the closest meteorological variable to solar radiation [57]. Thus, many scholars have established solar-radiation models based on air temperature. In addition, relative humidity and wind speed have also been used to improve the accuracy of solar radiation prediction [58,59]. Although the forecast data set was used in this study, similar results have been obtained, which means that the forecast data set and observation data have similar results. The most significant difference between the forecast data set and observation data lies in the forecast precision of different forecast factors. In general, the temperature has a very high forecast accuracy, but the relative humidity and wind-speed forecast accuracy are low, a fact mainly caused by two data mismatches. That is to say, the forecast data is the average of a large area, while the relative humidity and wind speed observed by the weather station is a minimal point value. We found that, in the four stations of this study, the model's accuracy with temperature factor is generally better than that of wind speed and relative humidity, and the prediction performance of relative humidity and wind speed of GEFSv12 needs to be improved.

## 5. Conclusions

Accurate forecasting of solar radiation (Rs) is significant to photovoltaic power generation and agricultural management. For the first time, this study evaluated and improved the capability of the newly released National Centers for Environmental Prediction Global Ensemble Forecast System version 12 (NECP GEFSv12) for short-term forecasting of Rs. To achieve this goal, a new coupling model based on the bat algorithm (BA) and kernel-based nonlinear extension of Arps decline (KNEA) was established. The data used four solar-radiation stations in Xinjiang, China as the benchmark. The new model was also compared with two empirical statistical methods (quantile mapping and Equiratio cumulative distribution function matching) with five machine-learning methods, e.g., support vector machine (SVM), XGBoost, KNEA, BA-SVM, BA-XGBoost. The results show that the accuracy of forecasting Rs from all of the models decreases from 1 d to 3 d ahead. Compared with the GEFS raw Rs data over the four stations, the RMSE and MAE of the QM and EDCDFm models decreased by 20% and 15%, respectively. In addition, the BA-KNEA model was superior to the GEFSv12 raw Rs data and other post-processing methods, with $R^2$ = 0.782–0.829, RMSE= 3.240–3.685 MJ $m^{-2}$ $d^{-1}$, MAE = 2.465–2.799 MJ $m^{-2}$ $d^{-1}$, NRMSE = 0.152–0.173.

**Author Contributions:** Conceptualization, L.W.; methodology, L.W. and F.L.; software, S.W.; validation, G.D.; formal analysis, G.D.; investigation, G.D.; resources, data curation, S.W.; writing—original draft preparation, L.W. and G.D.; writing—review and editing, G.D. and L.W.; supervision, F.L.; project administration, Y.W.; funding acquisition, S.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was jointly supported by the National Natural Science Foundation of China (No. 51879226, 51709143) and Jiangxi Natural Science Foundation of China (No. 20181BBG78078). The APC was funded by Jiangxi Natural Science Foundation of China (No. 20181BBG78078).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** Thanks to the National Meteorological Information Center of China Meteorological Administration for offering the meteorological data.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Liu, P.; Tong, X.; Zhang, J.; Meng, P.; Li, J.; Zheng, J. Estimation of half-hourly diffuse solar radiation over a mixed plantation in north China. *Renew. Energy* **2020**, *149*, 1360–1369. [CrossRef]
2. Demircan, C.; Bayrakçı, H.C.; Keçebaş, A. Machine learning-based improvement of empiric models for an accurate estimating process of global solar radiation. *Sustain. Energy Technol. Assess.* **2020**, *37*, 100574. [CrossRef]
3. Chang, K.; Zhang, Q. Improvement of the hourly global solar model and solar radiation for air-conditioning design in China. *Renew. Energy* **2019**, *138*, 1232–1238. [CrossRef]
4. Zang, H.; Cheng, L.; Ding, T.; Cheung, K.W.; Wang, M.; Wei, Z.; Sun, G. Application of functional deep belief network for estimating daily global solar radiation: A case study in China. *Energy* **2019**, *191*, 116502. [CrossRef]
5. Rehman, S.; Mohandes, M. Artificial neural network estimation of global solar radiation using air temperature and relative humidity. *Energy Policy* **2008**, *36*, 571–576. [CrossRef]
6. Quej, V.H.; Almorox, J.; Arnaldo, J.A.; Saito, L. ANFIS, SVM and ANN soft-computing techniques to estimate daily global solar radiation in a warm sub-humid environment. *J. Atmos. Sol.-Terr. Phys.* **2017**, *155*, 62–70. [CrossRef]
7. Ghimire, S.; Deo, R.C.; Raj, N.; Mi, J. Deep solar radiation forecasting with convolutional neural network and long short-term memory network algorithms. *Appl. Energy* **2019**, *253*, 113541. [CrossRef]
8. Deo, R.C.; Şahin, M.; Adamowski, J.F.; Mi, J. Universally deployable extreme learning machines integrated with remotely sensed MODIS satellite predictors over Australia to forecast global solar radiation: A new approach. *Renew. Sustain. Energy Rev.* **2019**, *104*, 235–261. [CrossRef]
9. Hassan, M.A.; Khalil, A.; Kaseb, S.; Kassem, M.A. Exploring the potential of tree-based ensemble methods in solar radiation modeling. *Appl. Energy* **2017**, *203*, 897–916. [CrossRef]
10. Güçlü, Y.S.; Yeleğen, M.Ö.; Dabanlı, İ.; Şişman, E. Solar irradiation estimations and comparisons by ANFIS, Angström–Prescott and dependency models. *Sol. Energy* **2014**, *109*, 118–124. [CrossRef]
11. Mohammadi, K.; Shamshirband, S.; Kamsin, A.; Lai, P.C.; Mansor, Z. Identifying the most significant input parameters for predicting global solar radiation using an ANFIS selection procedure. *Renew. Sustain. Energy Rev.* **2016**, *63*, 423–434. [CrossRef]
12. Feng, Y.; Cui, N.; Chen, Y.; Gong, D.; Hu, X. Development of data-driven models for prediction of daily global horizontal irradiance in northwest China. *J. Clean. Prod.* **2019**, *223*, 136–146. [CrossRef]
13. Wu, L.; Huang, G.; Fan, J.; Zhang, F.; Wang, X.; Zeng, W. Potential of kernel-based nonlinear extension of Arps decline model and gradient boosting with categorical features support for predicting daily global solar radiation in humid regions. *Energy Convers. Manag.* **2019**, *183*, 280–295. [CrossRef]
14. Fan, J.; Wu, L.; Zhang, F.; Cai, H.; Wang, X.; Lu, X.; Xiang, Y. Evaluating the effect of air pollution on global and diffuse solar radiation prediction using support vector machine modeling based on sunshine duration and air temperature. *Renew. Sustain. Energy Rev.* **2018**, *94*, 732–747. [CrossRef]
15. Fan, J.; Wu, L.; Ma, X.; Zhou, H.; Zhang, F. Hybrid support vector machines with heuristic algorithms for prediction of daily diffuse solar radiation in air-polluted regions. *Renew. Energy* **2020**, *145*, 2034–2045. [CrossRef]
16. Belaid, S.; Mellit, A. Prediction of daily and mean monthly global solar radiation using support vector machine in an arid climate. *Energy Convers. Manag.* **2016**, *118*, 105–118. [CrossRef]
17. Urraca, R.; Martinez-de-Pison, E.; Sanz-Garcia, A.; Antonanzas, J.; Antonanzas-Torres, F. Estimation methods for global solar radiation: Case study evaluation of five different approaches in central Spain. *Renew. Sustain. Energy Rev.* **2017**, *77*, 1098–1113. [CrossRef]
18. Álvarez-Alvarado, J.M.; Ríos-Moreno, J.G.; Obregón-Biosca, S.A.; Ronquillo-Lomelí, G.; Ventura-Ramos, E.; Trejo-Perea, M. Hybrid techniques to predict solar radiation using support vector machine and search optimization algorithms: A review. *Appl. Sci.* **2021**, *11*, 1044. [CrossRef]
19. Dong, J.; Wu, L.; Liu, X.; Fan, C.; Leng, M.; Yang, Q. Simulation of daily diffuse solar radiation based on three machine learning models. *Comput. Model. Eng. Sci.* **2020**, *123*, 49–73. [CrossRef]
20. Feng, Y.; Hao, W.; Li, H.; Cui, N.; Gong, D.; Gao, L. Machine learning models to quantify and map daily global solar radiation and photovoltaic power. *Renew. Sustain. Energy Rev.* **2020**, *118*, 109393. [CrossRef]
21. Liu, Y.; Zhou, Y.; Chen, Y.; Wang, D.; Wang, Y.; Zhu, Y. Comparison of support vector machine and copula-based nonlinear quantile regression for estimating the daily diffuse solar radiation: A case study in China. *Renew. Energy* **2020**, *146*, 1101–1112. [CrossRef]
22. Qing, X.; Niu, Y. Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM. *Energy* **2018**, *148*, 461–468. [CrossRef]
23. Abdel-Nasser, M.; Mahmoud, K. Accurate photovoltaic power forecasting models using deep LSTM-RNN. *Neural Comput. Appl.* **2019**, *31*, 2727–2740. [CrossRef]

24. Huang, C.; Kuo, P. Multiple-input deep convolutional neural network model for short-term photovoltaic power forecasting. *IEEE Access* **2019**, *7*, 74822–74834. [CrossRef]
25. Kaba, K.; Sarıgül, M.; Avcı, M.; Kandırmaz, H.M. Estimation of daily global solar radiation using deep learning model. *Energy* **2018**, *162*, 126–135. [CrossRef]
26. Voyant, C.; Notton, G.; Kalogirou, S.; Nivet, M.; Paoli, C.; Motte, L.; Fouilloy, A. Machine learning methods for solar radiation forecasting: A review. *Renew. Energy* **2017**, *105*, 569–582. [CrossRef]
27. Sun, H.; Gui, D.; Yan, B.; Liu, Y.; Liao, W.; Zhu, Y.; Lu, C.; Zhao, N. Assessing the potential of random forest method for estimating solar radiation using air pollution index. *Energy Convers. Manag.* **2016**, *119*, 121–129. [CrossRef]
28. Ibrahim, I.A.; Khatib, T. A novel hybrid model for hourly global solar radiation prediction using random forests technique and firefly algorithm. *Energy Convers. Manag.* **2017**, *138*, 413–425. [CrossRef]
29. Prasad, R.; Ali, M.; Kwan, P.; Khan, H. Designing a multi-stage multivariate empirical mode decomposition coupled with ant colony optimization and random forest model to forecast monthly solar radiation. *Appl. Energy* **2019**, *236*, 778–792. [CrossRef]
30. Hamill, T.M.; Whitaker, J.S.; Shlyaeva, A.; Bates, G.; Fredrick, S.; Pegion, P.; Sinsky, E.; Zhu, Y.; Tallapragada, V.; Guan, H.; et al. The Reanalysis for the Global Ensemble Forecast System, Version 12. Monthly. *Weather Rev.* **2022**, *150*, 59–79. [CrossRef]
31. Fan, J.; Chen, B.; Wu, L.; Zhang, F.; Lu, X.; Xiang, Y. Evaluation and development of temperature-based empirical models for estimating daily global solar radiation in humid regions. *Energy* **2018**, *144*, 903–914. [CrossRef]
32. Zhou, X.; Zhu, Y.; Hou, D.; Fu, B.; Li, W.; Guan, H.; Sinsky, E.; Kolczynski, W.; Xue, X.; Luo, Y.; et al. The Development of the NCEP Global Ensemble Forecast System Version 12. *Weather Forecast.* **2022**, *37*, 727. [CrossRef]
33. Tallapragada, V. Recent updates to NCEP Global Modeling Systems: Implementation of FV3 based Global Forecast System (GFS v15. 1) and plans for implementation of Global Ensemble Forecast System (GEFSv12). In *AGU Fall Meeting Abstracts*; Astrophysics Data System: San Francisco, CA, USA, 2019; pp. A31C–A34C.
34. Lee, T.; Singh, V.P. *Statistical Downscaling for Hydrological and Environmental Applications*; CRC Press: Boca Raton, FL, USA, 2018. [CrossRef]
35. Maraun, D. Bias correction, quantile mapping, and downscaling: Revisiting the inflation issue. *J. Clim.* **2013**, *26*, 2137–2143. [CrossRef]
36. Guo, L.; Gao, Q.; Jiang, Z.; Li, L. Bias correction and projection of surface air temperature in LMDZ multiple simulation over central and eastern China. *Adv. Clim. Chang. Res.* **2018**, *9*, 81–92. [CrossRef]
37. Yu, Y.; Si, X.; Hu, C.; Zhang, J. A review of recurrent neural networks: LSTM cells and network architectures. *Neural Comput.* **2019**, *31*, 1235–1270. [CrossRef]
38. Yan, R.; Liao, J.; Yang, J.; Sun, W.; Nong, M.; Li, F. Multi-hour and multi-site air quality index forecasting in Beijing using CNN, LSTM, CNN-LSTM, and spatiotemporal clustering. *Expert Syst. Appl.* **2021**, *169*, 114513. [CrossRef]
39. Ao, C.; Zeng, W.; Wu, L.; Qian, L.; Srivastava, A.K.; Gaiser, T. Time-delayed machine learning models for estimating groundwater depth in the Hetao Irrigation District, China. *Agric. Water Manag.* **2021**, *255*, 107032. [CrossRef]
40. Vapnik, V.N. An overview of statistical learning theory. *IEEE Trans. Neural Netw.* **1999**, *10*, 988–999. [CrossRef]
41. Chen, T.; He, T.; Benesty, M.; Khotilovich, V.; Tang, Y.; Benesty, M.; Khotilovich, V.; Tang, Y.; Cho, H.; Chen, K. *Xgboost: Extreme Gradient Boosting*; R Package Vers. 0.4-2; Xgboost: Seattle, WA, USA, 2015; pp. 1–4.
42. Ma, X.; Liu, Z. Predicting the oil production using the novel multivariate nonlinear model based on Arps decline model and kernel method. *Neural Comput. Appl.* **2018**, *29*, 579–591. [CrossRef]
43. Lu, H.; Ma, X.; Huang, K.; Azimi, M. Prediction of offshore wind farm power using a novel two-stage model combining kernel-based nonlinear extension of the Arps decline model with a multi-objective grey wolf optimizer. *Renew. Sustain. Energy Rev.* **2020**, *127*, 109856. [CrossRef]
44. Yang, X.; He, X. Bat algorithm: Literature review and applications. *Int. J. Bio-Inspired Comput.* **2013**, *5*, 141–149. [CrossRef]
45. Cui, Y.; Jia, L.; Fan, W. Estimation of actual evapotranspiration and its components in an irrigated area by integrating the Shuttleworth-Wallace and surface temperature-vegetation index schemes using the particle swarm optimization algorithm. *Agric. For. Meteorol.* **2021**, *307*, 108488. [CrossRef]
46. Wang, D.; Tan, D.; Liu, L. Particle swarm optimization algorithm: An overview. *Soft Comput.* **2018**, *22*, 387–408. [CrossRef]
47. Erfani, S.M.; Rajasegarar, S.; Karunasekera, S.; Leckie, C. High-dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning. *Pattern Recognit.* **2016**, *58*, 121–134. [CrossRef]
48. Zang, H.; Liu, L.; Sun, L.; Cheng, L.; Wei, Z.; Sun, G. Short-term global horizontal irradiance forecasting based on a hybrid CNN-LSTM model with spatiotemporal correlations. *Renew. Energy* **2020**, *160*, 26–41. [CrossRef]
49. Shin, D.; Ha, E.; Kim, T.; Kim, C. Short-term photovoltaic power generation predicting by input/output structure of weather forecast using deep learning. *Soft Comput.* **2021**, *25*, 771–783. [CrossRef]
50. Hu, M.; Zhao, B.; Ao, X.; Cao, J.; Wang, Q.; Riffat, S.; Su, Y.; Pei, G. Applications of radiative sky cooling in solar energy systems: Progress, challenges, and prospects. *Renew. Sustain. Energy Rev.* **2022**, *160*, 112304. [CrossRef]
51. Feng, Y.; Zhang, X.; Jia, Y.; Cui, N.; Hao, W.; Li, H.; Gong, D. High-resolution assessment of solar radiation and energy potential in China. *Energy Convers. Manag.* **2021**, *240*, 114265. [CrossRef]
52. De Araujo, J.M.S. Performance comparison of solar radiation forecasting between WRF and LSTM in Gifu, Japan. *Environ. Res. Commun.* **2020**, *2*, 045002. [CrossRef]

53. Zhou, Y.; Liu, Y.; Wang, D.; Liu, X.; Wang, Y. A review on global solar radiation prediction with machine learning models in a comprehensive perspective. *Energy Convers. Manag.* **2021**, *235*, 113960. [CrossRef]
54. Qiu, R.; Li, L.; Wu, L.; Agathokleous, E.; Liu, C.; Zhang, B.; Luo, Y.; Sun, S. Modeling daily global solar radiation using only temperature data: Past, development, and future. *Renew. Sustain. Energy Rev.* **2022**, *163*, 112511. [CrossRef]
55. Makade, R.G.; Chakrabarti, S.; Jamil, B. Development of global solar radiation models: A comprehensive review and statistical analysis for Indian regions. *J. Clean. Prod.* **2021**, *293*, 126208. [CrossRef]
56. Tao, H.; Ewees, A.A.; Al-Sulttani, A.O.; Beyaztas, U.; Hameed, M.M.; Salih, S.Q.; Armanuos, A.M.; Al-Ansari, N.; Voyant, C.; Shahid, S.; et al. Global solar radiation prediction over North Dakota using air temperature: Development of novel hybrid intelligence model. *Energy Rep.* **2021**, *7*, 136–157. [CrossRef]
57. Zhang, Y.; Cui, N.; Feng, Y.; Gong, D.; Hu, X. Comparison of BP, PSO-BP and statistical models for predicting daily global solar radiation in arid Northwest China. *Comput. Electron. Agric.* **2019**, *164*, 104905. [CrossRef]
58. Yadav, A.K.; Chandel, S.S. Solar radiation prediction using Artificial Neural Network techniques: A review. *Renew. Sustain. Energy Rev.* **2014**, *33*, 772–781. [CrossRef]
59. Fan, J.; Wang, X.; Wu, L.; Zhang, F.; Bai, H.; Lu, X.; Xiang, Y. New combined models for estimating daily global solar radiation based on sunshine duration in humid regions: A case study in South China. *Energy Convers. Manag.* **2018**, *156*, 618–625. [CrossRef]