Machine learning models for the estimation of monthly mean daily reference evapotranspiration based on cross-station and synthetic data

Lifeng Wu, Youwen Peng, Junliang Fan and Yicheng Wang

ABSTRACT

The estimation of reference evapotranspiration (ET_0) is important in hydrology research, irrigation scheduling design and water resources management. This study explored the capability of eight machine learning models, i.e., Artificial Neuron Network (ANN), Random Forest (RF), Gradient Boosting Decision Tree (GBDT), Extreme Gradient Boosting (XGBoost), Multivariate Adaptive Regression Spline (MARS), Support Vector Machine (SVM), Extreme Learning Machine and a novel Kernel-based Nonlinear Extension of Arps Decline (KNEA) Model, for modeling monthly mean daily ET_0 using only temperature data from local or cross stations. These machine learning models were also compared with the temperature-based Hargreaves–Samani equation. The results indicated that the estimation accuracy of these machine learning models differed in various scenarios. The tree-based models (RF, GBDT and XGBoost) exhibited higher estimation accuracy than the other models in the local application. When the station has only temperature data, the MARS and SVM models were slightly superior to the other models, while the ANN and HS models performed worse than the others. When there was no temperature data at the target station and the data from adjacent stations were used instead, MARS, SVM and KNEA were the suitable models. The results can provide a solution for ET_0 estimation in the absence of complete meteorological data.

Key words | cross-station, Kernel-based nonlinear extension of arps decline model, machine learning, reference evapotranspiration

Lifeng Wu
Youwen Peng
School of Hydraulic and Ecological Engineering,
Nanchang Institute of Technology,
Nanchang 330099,
China

Lifeng Wu
Junliang Fan (corresponding author)
Key Laboratory of Agricultural Soil and Water Engineering in Arid and Semiarid Areas of Ministry of Education, Northwest A&F University,
Yangling 712100,
China
E-mail: *nwwfjl@163.com*

Lifeng Wu
Yicheng Wang
State Key Laboratory of Simulation and Regulation of Water Cycle in River Basin,
China Institute of Water Resources and Hydropower Research, Beijing 100038,
China

Junliang Fan
College of Water Resources and Architectural Engineering,
Northwest A&F University,
Yangling 712100, Shaanxi,
China

INTRODUCTION

Evapotranspiration (ET) is the combination of two separate water loss processes: water evaporation from the soil and plant surfaces and plant transpiration by which water escapes from a plant's body to the ambient air in the form of steam through its stomata (Ali Ghorbani *et al.* 2018; Moazenzadeh *et al.* 2018). ET is one of the important components of hydrologic cycle. Reliable estimation of ET is the basis of developing precision irrigation system and improving water use efficiency. Although ET can be measured using eddy covariance, Bowen ratio system or lysimeters, their common problem is that they are expensive, time-consuming and require high professional knowledge, particularly in developing countries such as China. People usually use indirect methods to derive ET, that is, to use reference crop evapotranspiration (ET_0) and crop coefficient (K_c). According to the Food and Agriculture Organization of the United Nations (FAO) publication by Allen *et al.* (1998), ET_0 represents 'the evapotranspiration from an actively growing virtual vegetated surface that is 0.12 m tall, completely shading the ground, with adequate water supply, and for daily time-step the aerodynamic resistance is $208/u_2$ (wind speed at 2 meter) surface albedo is 0.23 and a bulk canopy

doi: 10.2166/nh.2019.060

resistance is 70 s m^{-1}. The most widely accepted methodology for ET_0 estimation is the FAO 56 Penman–Monteith (PMF 56) formula, and it has been a standard method to test the other methods (Fan *et al.* 2016).

The main drawback of the PMF 56 formula is that it needs many high-quality meteorological data, e.g. solar radiation or sunshine duration, air temperature, wind speed and relative humidity, whereas these data are often unavailable in developing countries. For instance, the costs of observing solar radiation are very high. There are more than 2,000 meteorological stations in China, but only less than 130 stations record solar radiation (Fan *et al.* 2019a, 2019b). In addition, wind speed is affected by the topographic feature and land use, and it is difficult to obtain representative wind speed on a large scale. For this reason, temperature-based ET_0 models are of great interest to researchers. There have been many temperature-based empirical models available for ET_0 estimation, e.g. the Thornthwaite model (Pereira & Pruitt 2004; Beguería *et al.* 2014), Hamon model (McCabe *et al.* 2015; Valipour 2015), Malmström model (Almorox *et al.* 2015; Quej *et al.* 2018), Hargreaves–Samani (HS) model (Luo *et al.* 2014; Pandey *et al.* 2014; Shiri *et al.* 2015; Xu *et al.* 2016; Cobaner *et al.* 2017; Morales-Salinas *et al.* 2017), Oudin model (Oudin *et al.* 2005; Zhao *et al.* 2013), Blaney–Criddle model (Heydari *et al.* 2015; Valipour 2015; Valipour *et al.* 2017), and Baier–Robertson model (Liu *et al.* 2016; Seiller & Anctil 2016; Martel *et al.* 2018).

Among these temperature-based models, the Hargreaves–Samani model has been widely used all over the world as a result of its simple structure and strong applicability. Hargreaves and Allen (2003) suggested that suitable ET_0 estimates could be obtained by the HS model for at least a five-day period, since the daily value was easily influenced by wind speed and cloud cover. However, ideal results on a daily scale were also reported. Raziei and Pereira (2013) evaluated the performance of the HS and FAO-PM temperature (PMT) models for the estimation of ET_0 at 40 weather stations in Iran. The results suggested that the HS and PMT models had similar estimation accuracy in modeling ET_0 in various climatic zones of Iran. Almorox *et al.* (2015) assessed more than 10 temperature-based models for estimating ET_0 at 4362 worldwide stations. In this study, the HS model provided the best accuracy in many climates, e.g. arid, semiarid, temperate, cold and polar. On the other hand, the Thornthwaite and McCloud models gave the worst average estimates in all climates. Quej *et al.* (2018) evaluated seven temperature-based models for estimating ET_0 in four cities of Mexico. They found that the HS model exhibited satisfactory accuracy (root mean square error (RMSE) = 0.74 mm d^{-1}), which was slightly worse than that of the PMT model (RMSE = 0.70 mm d^{-1}) in Yucatán Peninsula. Although the HS model has a good performance worldwide, model parameter calibration is a crucial prerequisite for local applications (Samani 2000). Fourteen general parametric models were established based on the geographical, temperature and wind speed information by Martí *et al.* (2015) in eastern Spain. Feng *et al.* (2017) calibrated the HS model based on the Bayesian method in the Sichuan basin of Southwest China. Other regions have also reported the calibration of the HS model parameters (Gavilán *et al.* 2006; Ravazzani *et al.* 2011; Shahidian *et al.* 2013; Heydari & Heydari 2014; Almorox & Grieser 2016; Cobaner *et al.* 2017; Shiri 2017; Valiantzas 2017).

In recent years, there has been more and more research focusing on the estimation and forecast of natural phenomenon (Yaseen *et al.* 2018, 2018c; Fan *et al.* 2018a; Ghorbani *et al.* 2018a, 2018b; Khosravi *et al.* 2018; Naganna *et al.* 2019; Xiao *et al.* 2019), including ET_0 estimation by using machine learning models, e.g. Artificial Neural Network (ANN), Fuzzy Logic, Gene Expression Programming (GEP), Multivariate Adaptive Regression Splines (MARS), Decision Tree (DT), Random Forests (RFs), Support Vector Machine (SVM), Extreme Learning Machine (ELM) and Adaptive Neuro-fuzzy Inference System (ANFIS). Trajkovic (2005) compared the radial basis function neural network (RBFNN) model and three temperature-based empirical models (PMT, HS and Thornthwaite) for estimating ET_0 at seven weather stations in Serbia. The results showed that the RBFNN model provided better ET_0 estimates than the other models at most stations. Luo *et al.* (2015) evaluated four ANN models for ET_0 prediction using forecasted temperature data. The results showed that the average values of RMSE ranged 0.87–1.36 mm d^{-1}, and the prediction accuracy of maximum temperature was lower than that of minimum temperature. Yassin *et al.* (2016) compared the ANN and GEP models for estimating ET_0 in Saudi Arabia. The results indicated that the ANN model performed slightly better than the GEP model under the same input

combination of meteorological data. Feng *et al.* (2016) compared three machine learning models and several empirical models for the estimation of ET_0 in the humid region of Southwest China, and the ELM and GANN models were recommended. Similar work has also been done in Iran (Mehdizadeh *et al.* 2017). It was found that the MARS and SVM models offered better ET_0 estimates than the GEP and empirical models. Mattar (2018) developed a GEP model for estimating ET_0 at 32 weather stations in Egypt. It was found that the GEP model had better estimation accuracy than the empirical models. Fan *et al.* (2018c) evaluated the M5 model tree (M5Tree), Gradient Boosting Decision Tree (GBDT), RF, XGBoost, SVM and ELM models for predicting daily ET_0 in different climates of China. They found that the ELM and SVM models performed slightly better than the XGBoost model in terms of estimation accuracy, while the XGBoost model had much less computational time than the ELM and SVM models.

In addition, machine learning models can be coupled with preprocessing or parameter optimization algorithms, and the hybrid models usually perform better than the traditional machine learning models (Feng *et al.* 2018; Yaseen *et al.* 2018a, 2018b; Wu *et al.* 2019). Tao *et al.* (2018a, 2018b) developed a coupled model based on the ANFIS model and firefly algorithm (FFA) for estimating ET_0 in Burkina Faso. The new ANFIS–FFA model ($R^2 = 0.97$, RMSE $= 0.24$ mm d^{-1} and mean absolute percent error (MAPE) $= 0.035$) was superior to the ANFIS model ($R^2 = 0.89$, RMSE $= 0.38$ mm d^{-1} and MAPE $= 0.037$). Shiri (2018) introduced a new hybrid model based on the RF model and wavelet transform (WT) to estimate ET_0 using air temperature and wind speed data. The results revealed that the new hybrid model improved the estimation accuracy of the RF model and was superior to the empirical models. The ANN and ELM models coupled with WT have also showed superiority to the ordinary ANN and ELM models (Kisi & Alizamir 2018).

After training and testing by local dataset, the application of machine learning models to other regions with similar climatic conditions may still have great uncertainty (Feng *et al.* 2019; Huang *et al.* 2019). To overcome this limitation, many scholars have tested the performance of machine learning models when using exogenous data (Martí *et al.* 2015; Landeras *et al.* 2018; Shiri 2019; Shiri *et al.* 2019). Martí and Gasque (2011) explored the use of continentality index to evaluate the station's climate characteristics. The object station, which was selected based on this characteristic, was used to develop the ANN model to accomplish cross-station strategy. In addition, another new approach based on the geographical inputs has also been reported by Martí and Zarzo (2012). Karimi *et al.* (2017) evaluated the performance of the GEP and SVM models for ET_0 estimation in the humid region of South Korea. The model was developed and tested at each location in the first scenario, and the results showed that the machine learning models had superiority to the empirical models. In the second scenario, ET_0 was modeled using data from nearby stations and the generalized heuristic model was developed for the studied stations. They found that both the GEP and SVM models could fulfill these tasks, where the GEP model slightly outperformed the SVM model. Shiri *et al.* (2014) developed ANFIS models based on weather data from Spain and found that the model could successfully estimate ET_0 in both the arid and humid regions of Iran. Kisi (2016) found that the estimation accuracy of the LSSVM, MARS and M5Tree models in ET_0 modeling differed in various cross-station scenarios. The MARS model outperformed the other models when local input data were not available. However, the M5Tree model performed better than the others when both local input and output data were missing. Feng *et al.* (2017) applied the RF and GRNN models to estimate ET_0 in both local and cross-station scenarios, and found that both models could estimate ET_0 accurately in the Sichuan Province of China. Sanikhani *et al.* (2019) evaluated six temperature-based machine learning models (GRNN, RBFNN, ANFIS-GP, ANFIS-SC, GEP and MLP) and the HS model for the estimation of ET_0 at two stations in Turkey. The results indicated that the machine learning models, except the MLP model, were superior to the HS model in the cross-station scenario.

Jiangxi Province is located in South China, which experiences a subtropical monsoon climate. This region is a major producing area for double-cropping rice and citrus fruits in China. Variation in seasonal precipitation distribution is obvious, resulting in the frequent existence of seasonal drought in this region. In 2018, for instance, the region was hit by a severe drought, which affected more than 200,000 ha areas and more than 3 million people, and caused direct economic losses of 240 million US dollars. Therefore, the reliable estimation of ET_0 is of crucial

significance for the rational utilization of agricultural water resources in this region. To the best knowledge of the authors, comprehensive comparison of various types of machine learning models for ET_0 estimation has been very minimal, especially their performances with limited temperature data in local and cross-station applications. Machine learning models have different precision performance in various regions. The most suitable model in Jiangxi Province has not been reported yet and this is the first time to compare various types of models for the estimation of ET_0 in this region. In addition, an improved version of kernel-based learning model, i.e., Kernel-based Nonlinear Extension of Arps decline (KNEA) model (Ma 2019), has been recently developed and successfully applied in many other fields (Ma & Liu 2018a). However, the KNEA model has not yet been tested in ET_0 studies. Therefore, this study aims to evaluate and compare the performance of eight temperature-based machine learning models, i.e. ANN, RF, GBDT, XGBoost, MARS, SVM, ELM and KNEA models for: (1) locally estimating monthly mean daily ET_0 at 15 stations in the Jiangxi Province of China using only temperature data, and compare their performance with the empirical HS model; (2) evaluating the developed models for estimating monthly mean daily ET_0 with data from four stations; and (3) evaluating the model performance for estimating monthly mean daily ET_0 using a new synthetic dataset (local extraterrestrial radiation data and temperature data from other stations).

MATERIALS AND METHODS

Case study and data description

Jiangxi Province, covering an area of 1.67×10^5 km^2, is a major producing area for double-cropping paddy in China and yields paddy rice of 20.4 billion kg y^{-1}. The study area has a subtropical humid climate with the mean annual rainfall ranging 1,341–1,943 mm, which is largely influenced by the East Asian monsoon (Fan *et al.* 2018a). About 15 billion m^3 y^{-1} of water resources have been used for irrigation in this region. However, nearly 2 billion m^3 y^{-1} water shortage exists as a result of unreasonable use of water resources and uneven distribution of seasonal rainfall. In this study, monthly maximum and minimum temperature data and extraterrestrial solar radiation from 15 meteorological stations in Jiangxi Province of China (Figure 1) were selected for testing the machine learning models and the empirical HS model in monthly ET_0 modeling. The meteorological data were examined and shared by the National Meteorological Information Center (NMIC) of China Meteorological Administration. The extraterrestrial solar radiation (R_a) data were estimated on the basis of geographical, seasonal and solar information (Quej *et al.* 2017). It can be seen from Table 1 that there was no significant variation in the meteorological variables between the training and testing periods at all stations. In addition, the temperature of Station 58506 was much lower than that of the other stations. However, the average annual ET_0 was slightly lower compared with the others due to higher elevation. The values of meteorological variables of the other stations (except Station 58506) are very similar, indicating that the air temperature and ET_0 data in this area had fewer variations. This makes it possible to develop general models for monthly ET_0 estimation in the whole region.

Machine learning models for estimating reference evapotranspiration

Gradient Boosting Decision Tree

The DT is one of the most widely used classification algorithms, which can be represented as multiple if-else rules. Decision tree is actually a method to divide the space into hyper planes. Each time the space is divided, the current space is divided into two parts, such as the decision tree, which makes each leaf node an intersecting region of space. After getting the above decision tree learning, when entering a classification samples instance for decision-making, we can divide the sample into a leaf node according to the two characteristics of the sample (x, y) values and classification results. This is the decision tree model of the classification process. The learning algorithm of decision tree has many subclasses, among which the ID3 algorithm, C4.5 and M5 model tree are the basic algorithms. The GBDT model is a hybridized algorithm that consists of an ensemble of decision trees. One single decision tree usually causes over-fitting issue, while the GBDT model is able to overcome this problem via integrating many weak decision trees

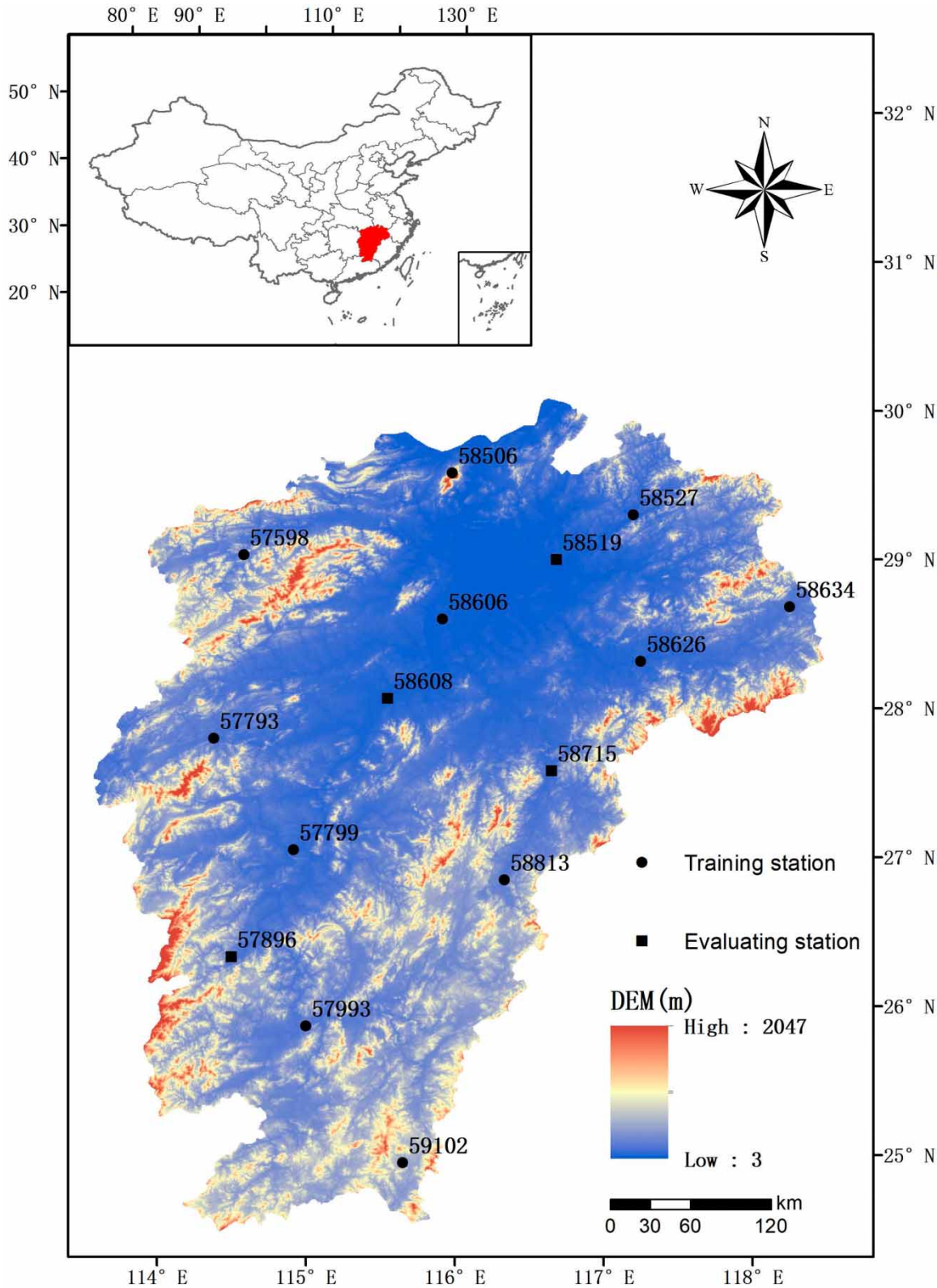

Figure 1 | Geographical locations of the 15 weather stations in the Jiangxi Province of China used in this study.

Table 1 | Geographic information, monthly mean daily values of sunshine duration (n), maximum and minimum air temperatures (T_{max} and T_{min}), relative humidity (RH), wind speed (U_2) and reference evapotranspiration (ET_0) during training (2001–2010) and testing (2011–2015, in brackets) for each of the 15 studied stations

ID	Longitude (°)	Latitude (°)	Elevation (m)	n (h)	T_{max} (°C)	T_{min} (°C)	RH(%)	U_2 (m s^{-1})	ET_0 (mm d^{-1})
57598	114.35	29.02	146.8	4.8 (4.5)	23.4 (23.1)	13.1 (13.5)	76.9 (80.4)	1.1 (1.1)	2.3 (2.2)
57793	114.23	27.48	131.3	4.0 (3.9)	22.8 (22.6)	14.5 (14.8)	79.1 (80.9)	1.9 (1.8)	2.3 (2.2)
57799	114.55	27.03	71.2	4.3 (4.0)	23.8 (23.5)	15.8 (15.9)	78.8 (78.5)	1.6 (1.5)	2.4 (2.3)
57896	114.30	26.20	126.1	4.6 (4.4)	24.3 (24.0)	15.7 (15.9)	75.1 (77.3)	1.7 (1.7)	2.6 (2.5)
57993	115.00	25.52	137.5	4.8 (4.8)	24.8 (24.8)	16.4 (16.8)	70.6 (74.5)	1.4 (1.5)	2.7 (2.7)
58506	115.59	29.35	1164.5	4.7 (4.3)	16.4 (16.4)	9.6 (9.5)	75.1 (78.2)	3.6 (3.5)	2.2 (2.0)
58519	116.41	29.00	40.1	4.9 (4.8)	22.7 (22.4)	15.4 (15.4)	73.0 (74.7)	2.0 (1.9)	2.6 (2.5)
58527	117.12	29.18	61.5	4.8 (4.6)	23.6 (23.3)	14.7 (14.7)	73.4 (74.2)	1.3 (1.1)	2.4 (2.3)
58606	115.55	28.36	46.9	5.3 (4.9)	22.6 (22.5)	15.6 (15.7)	71.8 (73.0)	1.9 (1.8)	2.7 (2.6)
58608	115.33	28.04	30.4	4.6 (4.2)	23.3 (23.1)	15.6 (15.6)	75.0 (73.4)	1.2 (1.2)	2.4 (2.4)
58626	117.15	28.19	60.8	4.5 (4.3)	23.9 (23.2)	15.8 (15.3)	74.4 (76.2)	1.4 (2.1)	2.5 (2.5)
58634	118.15	28.41	116.3	4.7 (4.5)	23.5 (23.1)	14.6 (14.7)	74.6 (75.6)	2.0 (2.0)	2.6 (2.5)
58715	116.39	27.35	80.8	4.7 (4.5)	23.4 (23.1)	15.1 (15.4)	78.2 (76.0)	2.6 (2.5)	2.6 (2.7)
58813	116.20	26.51	143.8	4.5 (4.1)	24.4 (24.1)	15.2 (15.3)	80.6 (79.2)	1.3 (1.4)	2.4 (2.3)
59102	115.39	24.57	303.9	4.6 (4.4)	25.0 (24.9)	15.4 (15.8)	77.3 (79.8)	1.2 (1.0)	2.4 (2.3)

with many leaf nodes. The GBDT model has many merits, such as the capability to identify nonlinear transformations, the capability to deal with a categorical variable, computational robustness and high scalability. GBDT had been used in web search (Mohan *et al.* 2011), subway ridership (Ding *et al.* 2016), global solar radiation (Fan *et al.* 2018b), pan evaporation (Lu *et al.* 2018) and ET_0 estimation (Fan *et al.* 2018a). More details can be found in Elith *et al.* (2008).

Extreme Gradient Boosting

The XGBoost model is proposed by Chen & Guestrin (2016), which is an improved version of Gradient Boosting Machines (GBMs) and in particular K Classification and Regression Trees (CART). This model is originated from the idea of 'boosting', which integrates all the predictions of a series of 'weak' learners to develop a 'strong' learner via an additive training process. The XGBoost model is supposed to prevent over-fitting issue and minimize the computational time. This is obtained by simplifying the objective functions that allow combining the predictive and regularization terms, while it maintains an optimal computational efficiency at the same time. Parallel calculations are also automatically executed for the functions in the XGBoost model in the training stage. More information about the XGBoost model refer to Chen & Guestrin (2016).

Kernel-based Nonlinear Extension of Arps Decline Model

KNEA is a newly nonlinear model initially proposed by Ma & Liu (2018a, 2018b) based on the Arps decline model (Ma *et al.* 2019a, 2019b) and kernel method (Vapnik 2013). KNEA can be described as:

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

where $f(x)$ is the output at this time and $f(x-1)$ is the output at the last step time. $u(x)$ is the factors which have effect on output, $g(u(x))$ can be interpreted as the relationship between $u(x)$ and $f(x)$. μ is a bias. From this model, we can see that the output of this time is the result of joint action between the output from last time-step and the influencing factors are at this time. The nonlinear function g is hard to determine and can be translated to:

$$g(u(x)) = \omega^T \varphi(u(x)) \tag{2}$$

This means mapping the original influence factors into the new space. Formula (2) can be thus written as:

$$f(x) = af(x-1) + \omega^T \varphi(u(x)) + \mu \tag{3}$$

Although we still cannot solve Equation (3), we can find a very small value so that the difference between the left and right of the equation is as small as possible:

$$e_x = f(x) - af(x-1) - \omega^T \varphi(u(x)) - \mu \tag{4}$$

$$\min_{a,\omega,e} \varsigma(a,\omega,e) = \frac{1}{2}a^2 + \frac{1}{2}\|\omega\|^2 + \frac{\gamma}{2}\sum_{x=2}^{n} e_x^2$$

$$\text{s.t. } f(x) = af(x-1) + \omega^T \varphi(u(x)) + \mu + e_x \tag{5}$$

where γ is called a regularization term, it can controls the smoothness of the model. Like SVM, this optimization problem can be used by the Lagrangian multiplier method:

$$\mathcal{L}(a,\omega,\mu,e,\lambda) = \varsigma(a,\omega,e) - \sum_{x=2}^{n} \lambda_x \{bf(x-1) + \omega^T \varphi(u(x)) + \mu + e_x - f(x)\} \tag{6}$$

where λ_x is the Lagrangian multiplier. The KKT conditions for optimality of the Lagrangian multiplier method are the following formulas:

$$\begin{cases} \dfrac{\partial \mathcal{L}}{\partial a} = 0 & a = \sum\limits_{x=2}^{n} \lambda_x f(x-1) \\ \dfrac{\partial \mathcal{L}}{\partial \omega} = 0 & \omega = \sum\limits_{x=2}^{n} \lambda_x \varphi(u(x)) \\ \dfrac{\partial \mathcal{L}}{\partial \mu} = 0 & \sum\limits_{x=2}^{n} \lambda_x = 0 \\ \dfrac{\partial \mathcal{L}}{\partial e_x} = 0 & e_x = \lambda_x \gamma^{-1} \\ \dfrac{\partial \mathcal{L}}{\partial \lambda_x} = 0 & f(x) = af(x-1) + \omega^T \varphi(u(x)) + \mu + e_x \end{cases} \tag{7}$$

$$\begin{pmatrix} 0 & 1_{n-1}^T \\ 1_{n-1} & \Omega + Q + \dfrac{1}{\gamma} I_{n-1} \end{pmatrix} \begin{pmatrix} \mu \\ \lambda \end{pmatrix} = \begin{pmatrix} 0 \\ x_{2|n} \end{pmatrix} \tag{8}$$

where

$$I_{n-1} = [1, 1, \ldots, 1]_{n-1}^T,$$

$$\lambda_{n-1} = [\lambda_1, \lambda_2, \ldots, \lambda_{n-1}]_{n-1}^T,$$

$$x_{2|n} = [x(2), x(3), \cdots, x(n)]^T,$$

$$\Omega_{ij} = \varphi(u(i)) \times \varphi(u(j)) = K(u(i), u(j)),$$

in which I_{n-1} is $n-1$ dimensional identity matrix with all the diagonal elements to be 1 and others to be 0. λ, μ and a can be obtained by Equation (9). The Ω_{ij} can be employed a kernel function $K(\cdot,\cdot)$ which satisfies the Mercer's theorem, and a RBF-type kernel function was selected in this study. More details about KNEA can be found in Ma & Liu (2018a).

$$K(u(i), u(j)) = \exp\left[\frac{-(u(i) - u(j))^2}{2\sigma^2}\right] \tag{9}$$

In addition, Artificial Neuron Network (ANN), Support Vector Machine (SVM), RF, Multivariate Adaptive Regression Spline (MARS) and Extreme Learning Machine (ELM) were also used in this study, and the details of these models can be found in Friedman (1991), Breiman (2001), Huang *et al.* (2006) and Vapnik (2013).

FAO 56 Penman–Monteith

The PMF 56 equation suggested by Allen *et al.* (1998) was used to calculate monthly mean daily ET_0 (mm d^{-1}) and provide the reference data for testing of the empirical and machine learning models in this paper, which can be calculated as:

$$ET_0 = \frac{0.408(R_n - G) + \gamma \dfrac{900}{T_a + 273} U(e_s - e_a)}{\Delta + \gamma(1 + 0.34U)} \tag{10}$$

where R_n is the net radiation at the crop surface, which is usually calculated by global solar radiation (R_s); G is the soil heat flux density; T_a is the mean daily air temperature at 2 m height, calculated as the mean of maximum (T_{max}) and minimum (T_{min}) air temperature; U is the wind speed at 2 m height; e_s and e_a are the saturation and actual vapor pressure; Δ is the slope of vapor pressure curve; γ is the air psychrometric constant. As in daily time-step in this study, G can be

neglected. The details of the PMF 56 equation can be found in Allen *et al.* (1998).

Hargreaves–Samani model

The temperature-based HS equation proposed by Hargreaves & Samani (1985) was used to estimate monthly mean daily ET_0 when only air temperature data are available:

$$ET_{0-HS} = 0.0023R_a(T_a + 17.8)\sqrt{T_{max} - T_{min}} \quad (11)$$

As mentioned above, this model will cause underestimation or overestimation of ET_0 without parameter calibration. In this study, the model can be written as follows:

$$ET_{0-HS} = aR_a(T_a + b)(T_{max} - T_{min})^c \quad (12)$$

where a, b and c are empirical coefficients. In this study, the PMF 56 equation with air temperature data only (PMT) has been tested for the estimation of monthly mean daily ET_0. However, it has been found to be inferior to the HS equation.

Model scenarios

In the field of agricultural irrigation management, it is of great significance for decision-makers and planners to obtain the information of ET_0. In this study, eight machine learning models as well as the HS empirical model were developed and applied by using the temperature-based general model for the estimation of monthly mean daily ET_0 in the Poyang Lake Region of Jiangxi Province. The obtained results of the machine learning models were also compared with those estimated by the standard PMF 56 equation. Firstly, a general model for estimating ET_0 was established using data during 2001–2010 from 11 meteorological stations in the Poyang Lake Region (Figure 2). Secondly, the established model was tested in three cases: (1) comparing the eight machine learning models and the HS empirical model for the estimation of monthly mean daily ET_0 of the 11 stations using data from 2011 to 2015; (2) investigating the same predictive model and comparing their performance with the HS model based on input and output data from the other four stations (ID: 57896, 58509, 58608 and 58715) in the same region; and (3) investigating the same predictive model and comparing their performance with the HS model based on temperature data from the four neighboring stations (ID: 577793, 58606, 58813 and 58527) of the four target stations (ID: 57896, 58509, 58608 and 58715) and extraterrestrial radiation data from the four target stations (ID: 57896, 58509, 58608 and 58715), respectively. The second and third cases will be useful for regions lack of temperature data or with no local data at all. The coefficients of the empirical models were attained by the least-squares fitting method, while the parameters of machine learning models were optimized by the grid search technique.

Statistical indicators

Four commonly used comparison statistics were employed to evaluate the proposed models in this study, including RMSE (Huffman 1997), R^2 (Hsu & Chen 1996), MBE and NRMSE (Fan *et al.* 2019a). RMSE and NRMSE can reflect the overall estimation accuracy of the models. R^2 indicates how much percentage of the data can be interpreted by the model, but there may also be a tendency to overestimate or underestimate. MBE is an indicator that reflects the overall overestimation or underestimation of the model. The four statistical indicators can be expressed as follows:

$$R^2 = \frac{\left[\sum_{i=1}^{n}(Y_{i,m} - \bar{Y}_{i,m})(Y_{i,e} - \bar{Y}_{i,e})\right]^2}{\sum_{i=1}^{n}(Y_{i,m} - \overline{Y}_{i,m})^2 \sum_{i=1}^{n}(Y_{i,e} - \overline{Y}_{i,e})^2} \quad (13)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_{i,m} - Y_{i,e})^2} \quad (14)$$

$$MBE = \frac{1}{n}\sum_{i=1}^{n}(Y_{i,m} - Y_{i,e}) \quad (15)$$

$$NRMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_{i,m} - Y_{i,e})^2}\ /\bar{Y}_{i,m} \times 100\% \quad (16)$$

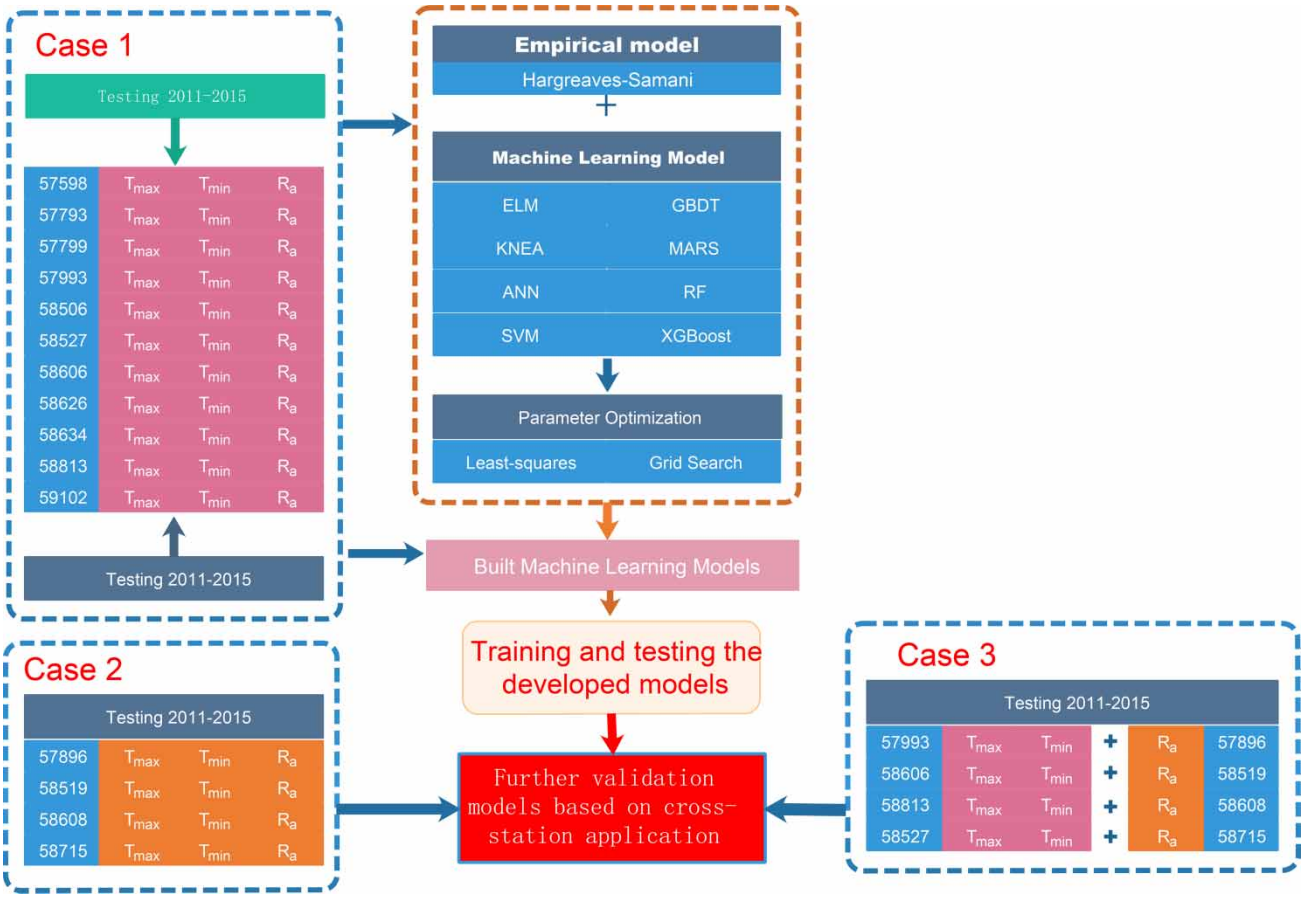

Figure 2 | Simple flowchart of the proposed methodology in the present study.

where $Y_{i,m}$, $Y_{i,e}$, $\overline{Y}_{i,m}$ and $\overline{Y}_{i,e}$ are the measured, estimated, mean of measured and mean of estimated ET_0 by the PMF 56 model, respectively; n is the number of observations. Higher R^2 values indicate high simulation accuracy, whereas the lower absolute values of RMSE, MBE and NRMSE suggest better model performance. Considering the requirements of the MLP and KNEA models, the raw meteorological data were normalized between 0 and 1 as follows:

$$z_n = \frac{z_i - z_{\min}}{z_{\max} - z_{\min}} \tag{17}$$

where z_n and z_i represent the moralized and raw training and testing data; z_{max} and z_{min} are the minimum and maximum of the training and testing data.

RESULTS AND DISCUSSION

Case 1

The comparison of the eight machine learning models as well as the HS model for estimating monthly mean daily ET_0 at the 11 stations in the Poyang Lake Region was performed. The statistical summary during training and testing are presented in Table 2. In general, MBE values were less than 0.05 mm d^{-1} during the training and testing periods. It means that there is no overall overestimation or underestimation by all the machine learning and empirical models. The tree-based models (RF, GBDT and XGBoost) had higher estimation accuracy during the testing stage. RMSE values of the RF, GBDT and XGBoost models were 0.276, 0.281 and 0.269 mm d^{-1} during testing. NRMSE

Table 2 | Statistics of the machine learning and empirical models for estimating ET_0 during training (2001–2010) and testing (2011–2015)

	Training				Testing			
Model	**RMSE (mm d^{-1})**	**R^2**	**NRMSE**	**MBE (mm d^{-1})**	**RMSE (mm d^{-1})**	**R^2**	**NRMSE**	**MBE (mm d^{-1})**
ELM	0.232	0.956	0.095	0.001	0.292	0.929	0.123	−0.024
GBDT	0.233	0.958	0.096	0.000	0.281	0.937	0.119	−0.047
KNEA	0.303	0.927	0.124	0.000	0.288	0.931	0.121	−0.024
MARS	0.308	0.924	0.126	0.000	0.295	0.929	0.124	−0.044
ANN	0.328	0.916	0.178	0.008	0.331	0.910	0.187	−0.032
RF	**0.130**	**0.986**	**0.053**	0.000	0.276	0.939	0.116	−0.046
SVM	0.308	0.925	0.126	0.033	0.294	0.929	0.124	−0.010
XGBoost	0.199	0.968	0.081	0.000	**0.269**	**0.941**	**0.113**	−0.040
HS	0.418	0.863	0.171	0.021	0.446	0.839	0.188	−0.016

Note: The best statistical indicators among the models are marked in bold.

values of the RF, GBDT and XGBoost models were 0.116, 0.119 and 0.113 during the testing stage. The RMSE values of the kernel-based models (ELM, KNEA and SVM) and the MARS model were close to each other, which were 2.5–9.3% higher than those of the tree-based models. The ANN model performed worst among machine learning models during the testing period. Compared with the XGBoost model, RMSE was increased by 23% during testing. However, the accuracy of the ANN model was significantly higher than that of the HS model, with the RMSE and NRMSE values of 0.446 mm d^{-1} and 0.188 during testing, respectively. It is clear that the worst model (HS) can still produce results that are suitable for estimating monthly mean daily ET_0 in this region. Overall, high estimation accuracy can be obtained by established models using only monthly mean daily maximum and minimum temperatures. This is because the global solar radiation in this area has a good relationship with daily maximum and minimum temperatures (Fan *et al.* 2018b). Also, the relative humidity is very high over the year and the influence of wind is not as obvious as that in the arid areas. Thus, the information most closely related to ET_0 can be described by temperature data alone. Similar results were also revealed in Southern China (Feng *et al.* 2017).

Figures 3 and 4 display the scatter plots of the PM-56 ET_0 and those estimated by the machine learning models and the HS empirical model during the training and testing periods, respectively. It is clear from the figure that all the nine models had passed the significance test ($P < 0.0001$). However, scatter plots of different models showed various distributions. The RF model ($R^2 = 0.987$) gave the less discrete points during the training period. The scatter distribution of the ELM ($R^2 = 0.987$), GBDT ($R^2 = 0.987$) and XGBoost ($R^2 = 0.987$) models were very close to each other during training. The KNEA, MARS, SVM and ANN models had more discrete distribution during training and they were close to each other. The HS model showed a serious underestimation when the PMF56 $ET_0 > 5$ mm d^{-1} during the training period. The ANN model produced more scattered estimates than the other machine learning models and the other eight machine learning models had a similar distribution of scatter points during testing. The scatter distribution was similar during testing and training for the HS model. This was not because the extreme values of the data were different between the two periods, but because the model itself did not capture useful information of temperature. In other words, diurnal temperature range and average temperature were not enough to describe the complex nonlinear relationship between temperature and ET_0.

To evaluate the balance of different machine models and the empirical model, the radar chart of RMSE at the 11 stations is presented in Figure 5. It can be clearly seen that the RMSE of the HS model was higher than the machine learning models at each site. The main reason for the low accuracy of the ANN model was that the stations in the south and west had larger errors than the other

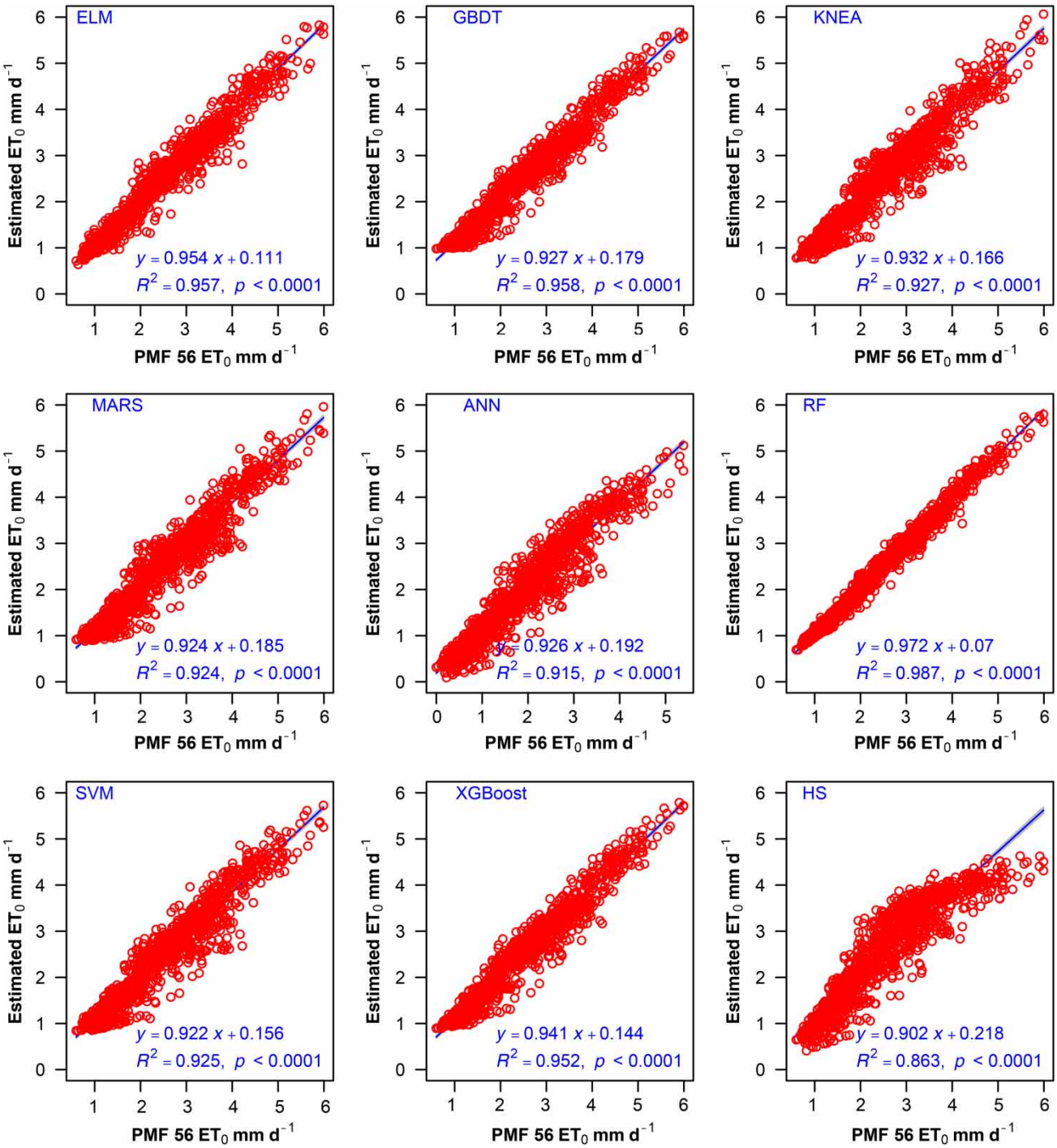

Figure 3 | Scatter plots of the PMF 56 ET_0 and those estimated by the machine learning models and the HS empirical model during the training period.

machine learning models. The accuracy of the SVM model was affected by the large errors of the three stations (ID: 57993, 58506 and 58527) in the north. The RF model ranked first at stations 59102 and 58813, but exhibited the moderate performance at the other stations. The GBDT and XGBoost models were very stable at each station and

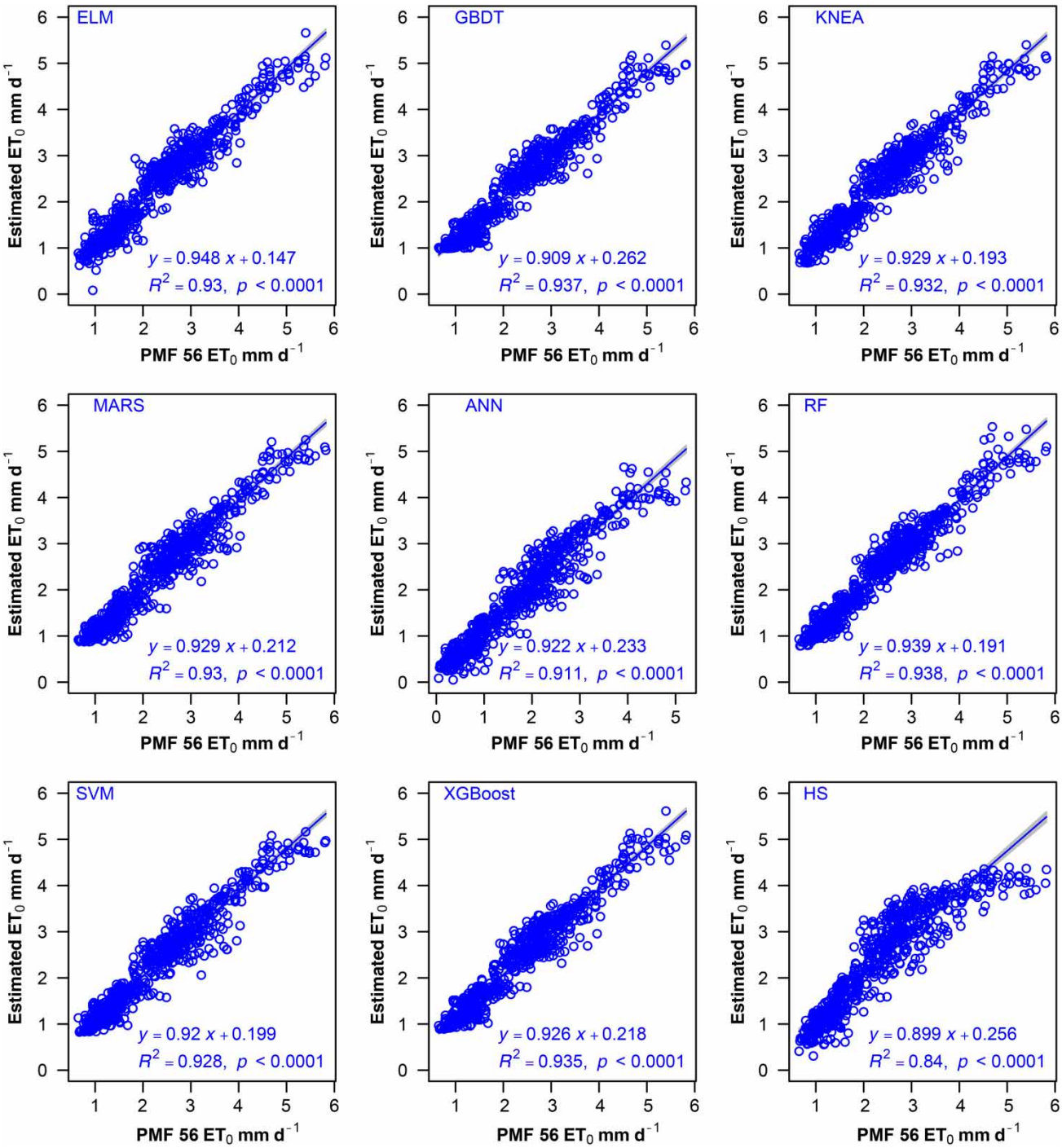

Figure 4 | Scatter plots of the PMF 56 ET_0 and those estimated by the machine learning models and the HS empirical model during the testing period.

ranked in the middle position. The above showed that different datasets had different impacts due to the various principles of model construction. However, due to the natural classification ability of tree-based models, different datasets can be converted into different decision trees, so the estimation accuracy of these models was higher.

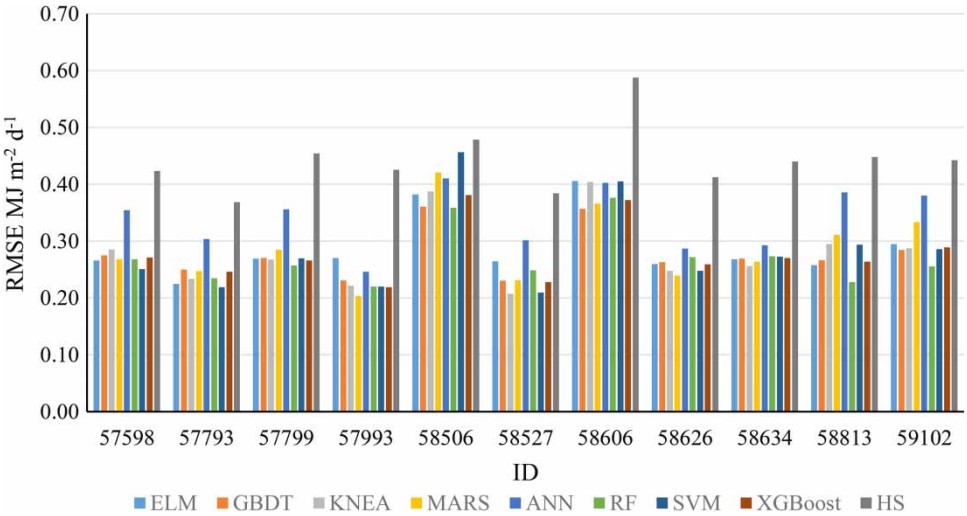

Figure 5 | Bar plot of RMSE values at the 11 stations.

Case 2

The model established in the previous section can be applied in the areas where only temperature observations are available. The application potential of different models in this case was further assessed, in which four stations with independent datasets (ID: 57896, 58519, 58608 and 58715) were used to evaluate the performance of the nine models. To clearly see the rank of the statistical results, the top three models were highlighted with red, green and blue colours. The mean values of the statistical results of the four stations are also listed in Table 3. As seen from the table, different models behaved differently at various stations. Taking Station 57896 as an example, the ranks of the studied models were: ELM> SVM > XGBoost > MARS > KNEA > GBDT > RF > ANN > HS. However, the SVM, XGBoost, MARS and KNEA models were close to each other. The ANN and HS models were worse than the other models, with the increase in RMSE by 10.5–37.1% and 35.4–67.9% at Station 57896 compared with the other models. From the statistical results of the average value of the four stations, the MARS and SVM models performed slightly superior to the other models, while the ANN and HS models performed worse than the others. It can be seen that since the variations of temperatures and ET_0 are small at each station, it is feasible to develop general models for the estimation of ET_0 in this region.

Figure 6 presents the scatter plots of the PM-56 ET_0 and those estimated by the same predictive machine learning models and the HS empirical model at the four stations. It is clear from the figure that all of the nine models had pass the significance test ($P < 0.0001$). All the machine learning models except the ANN model displayed relatively small scatter distribution. The HS model showed underestimation of monthly mean daily ET_0 to some extent when $ET_0 < 1.5$ mm d^{-1} or >5 mm d^{-1}.

Case 3

When a site lacks the basic temperature observation data, it can be replaced with temperature data from the other stations, which is commonly referred to as 'cross-station application'. In this section, it was supposed that there were four stations having no maximum and minimum temperature data, but only calculated extraterrestrial radiation data. The temperature data from the nearest station were used to replace the missing temperature data for each of these stations. In short, temperature data from stations 57896, 58519, 58813 and 58715 were replaced with those from stations 57993, 58527, 58606 and 58608, respectively. The statistical results are shown in Table 4. The GBDT model performed best at Station 57896, with RMSE 11.8–65.8% less than the other models. The KNEA, ELM, MARS, SVM and XGBoost models were close to each

Table 3 | Statistics of the machine learning and empirical models for estimating ET_0 during the testing period (2011–2015) using data from stations 57589, 58519, 58608 and 58715

ID	Model	RMSE (mm d^{-1})	R^2	NRMSE	MBE (mm d^{-1})
57896	ELM	0.237	0.968	0.093	0.102
	GBDT	0.263	0.962	0.103	0.033
	KNEA	0.246	0.96	0.096	0.033
	MARS	0.246	0.96	0.095	0.017
	ANN	0.325	0.945	0.128	0.091
	RF	0.294	0.945	0.116	0.042
	SVM	0.244	0.962	0.096	0.044
	XGBoost	0.244	0.962	0.096	0.048
	HS	0.398	0.891	0.154	0.031
58519	ELM	0.367	0.906	0.146	0.052
	GBDT	0.314	0.949	0.125	0.134
	KNEA	0.337	0.958	0.133	0.22
	MARS	0.294	0.958	0.116	0.148
	ANN	0.434	0.939	0.173	0.263
	RF	0.34	0.931	0.135	0.134
	SVM	0.332	0.962	0.132	0.215
	XGBoost	0.335	0.947	0.133	0.177
	HS	0.512	0.87	0.202	0.264
58608	ELM	0.282	0.955	0.12	− 0.126
	GBDT	0.242	0.968	0.103	− 0.119
	KNEA	0.237	0.966	0.099	− 0.093
	MARS	0.245	0.968	0.103	− 0.127
	ANN	0.257	0.955	0.109	− 0.017
	RF	0.249	0.964	0.106	− 0.12
	SVM	0.216	0.97	0.092	− 0.078
	XGBoost	0.236	0.964	0.1	− 0.091
	HS	0.381	0.893	0.16	− 0.018
58715	ELM	0.399	0.953	0.15	0.265
	GBDT	0.427	0.951	0.161	0.263
	KNEA	0.423	0.953	0.158	0.265
	MARS	0.392	0.951	0.146	0.224
	ANN	0.507	0.931	0.191	0.307
	RF	0.43	0.953	0.162	0.295
	SVM	0.417	0.953	0.157	0.264
	XGBoost	0.445	0.955	0.168	0.307
	HS	0.545	0.865	0.203	0.261
All	ELM	0.321	0.946	0.127	0.073
	GBDT	0.312	0.958	0.123	0.078
	KNEA	0.311	0.959	0.122	0.106
	MARS	0.294	0.959	0.115	0.066
	ANN	0.381	0.943	0.150	0.161
	RF	0.328	0.948	0.130	0.088
	SVM	0.302	0.962	0.119	0.111
	XGBoost	0.315	0.957	0.124	0.110
	HS	0.459	0.880	0.180	0.135

Note: the top three ranked models were highlighted in red, green and blue, respectively. Please refer to the online version of this paper to see this table in colour: https://dx.doi.org/10.2166/nh.2019.060.

other, while the RF, ANN and HS models were obviously not as good as these five models. However, the MARS, KNEA and SVM models ranked the top three models at Stations 58519, 58608 and 58715. On this basis, the MARS, SVM and KNEA models performed superior to the other models at all the four stations. Figure 7 presents the scatter plots of the PMF56 ET_0 and those estimated by the same predictive machine learning models and the HS empirical model performed at the other four stations in the cross-station applications. It is clear from the figure that all the nine models have passed the significance test ($P < 0.0001$). The scatter distribution of each model was not different from the performance of the models in the previous section. This indicates that it is feasible to use the adjacent meteorological data when local data are missing.

Kisi (2016) found that the LSSVM model was superior to the MARS model when local data were available, but the MARS model performed better than the LSSVM model when cross-station data were used. Similar results have been reported by Karimi *et al.* (2017) who found that the GEP model outperformed the SVM model in cross-station scenarios. In this study, the tree-based models performed better than the other models in local applications, while the MARS, SVM and KNEA models offered better ET_0 estimates than the others in the absence of local temperature data. This can be due to the differences in the dataset and the characteristics of various models. The tree-based models use greedy algorithms to explain every point as far as possible, but the dataset inevitably contain noise, which results in the over-fitting of the model to some extent. In addition, the tree-based models use many

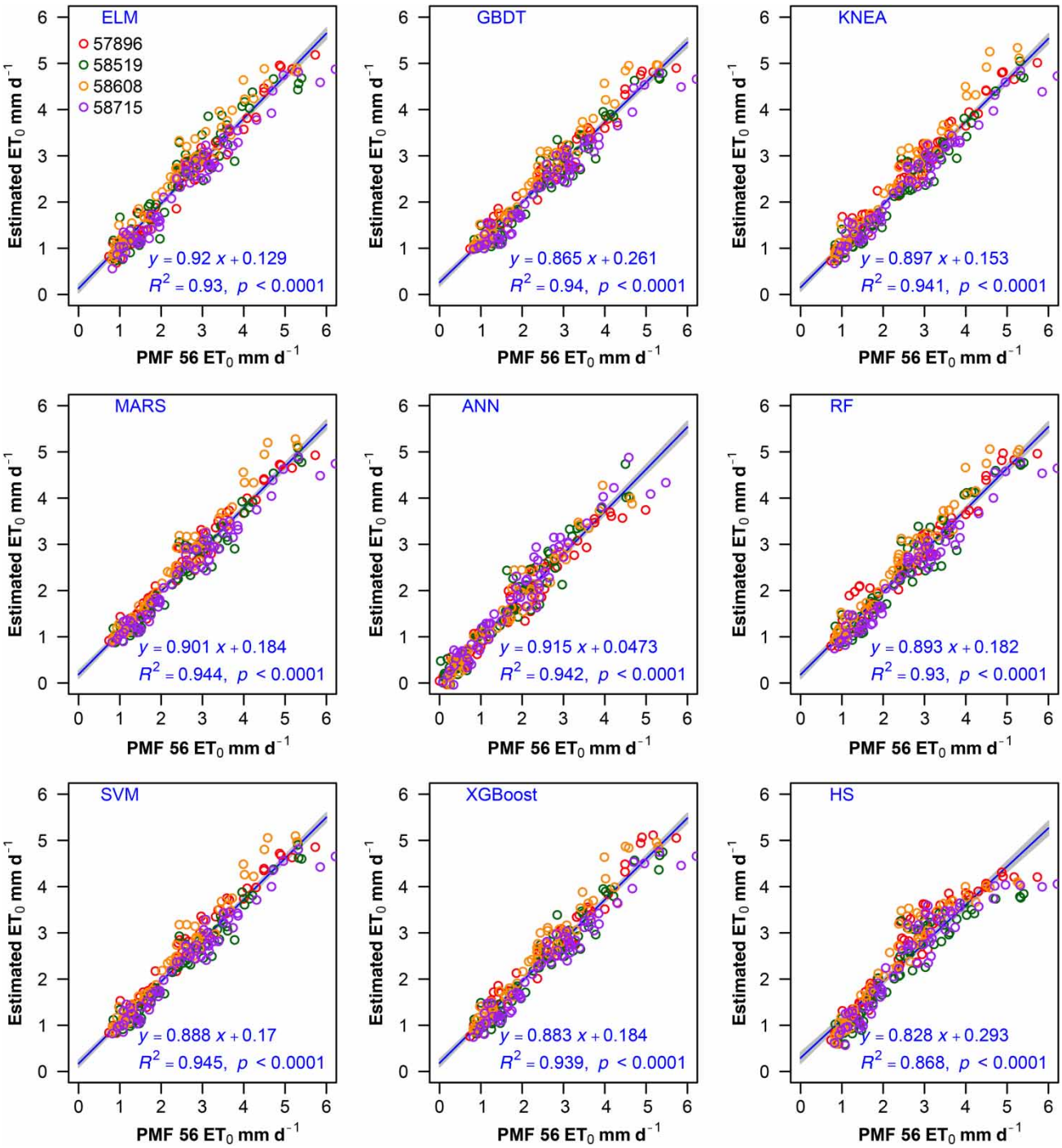

Figure 6 | Scatter plots of the PMF 56 ET_0 and those estimated by the same predictive machine learning models and the HS empirical model performed at the other four stations.

weak classifiers to establish a sub-model (one weak classifier) for small samples independently. The weight of this sub-model is much higher than that of other models, which can subtly obtain some useful information when localizing the model. However, this sub-model model may not be applicable in other regions, and it may also cause the over-fitting of the model when cross-station data are applied. The inspiration of the MARS model also comes

Table 4 | Statistics of the machine learning and empirical models for estimating ET_0 during the testing period (2011–2015) using new synthetic data (local extraterrestrial radiation data and temperature data from other station)

ID	Model	RMSE (mm d^{-1})	R^2	NRMSE	MBE (mm d^{-1})
57896	ELM	0.305	0.939	0.12	− 0.009
	GBDT	0.263	0.962	0.103	0.032
	KNEA	0.294	0.941	0.114	− 0.056
	MARS	0.302	0.941	0.117	− 0.086
	ANN	0.348	0.92	0.137	− 0.059
	RF	0.332	0.925	0.13	− 0.042
	SVM	0.301	0.939	0.118	− 0.047
	XGBoost	0.303	0.939	0.119	− 0.066
	HS	0.436	0.869	0.169	− 0.002
58519	ELM	0.308	0.945	0.123	0.08
	GBDT	0.282	0.955	0.112	0.077
	KNEA	0.272	0.96	0.107	0.096
	MARS	0.258	0.962	0.102	0.064
	ANN	0.313	0.941	0.125	0.087
	RF	0.287	0.947	0.114	0.076
	SVM	0.267	0.964	0.106	0.103
	XGBoost	0.273	0.956	0.109	0.075
	HS	0.458	0.857	0.181	0.087
58608	ELM	0.323	0.929	0.137	− 0.083
	GBDT	0.257	0.956	0.109	− 0.033
	KNEA	0.239	0.958	0.101	0.029
	MARS	0.226	0.962	0.095	− 0.035
	ANN	0.273	0.949	0.116	0.018
	RF	0.27	0.949	0.115	− 0.058
	SVM	0.221	0.966	0.094	0.026
	XGBoost	0.256	0.955	0.109	− 0.017
	HS	0.404	0.887	0.17	0.088
58715	ELM	0.416	0.925	0.157	0.185
	GBDT	0.434	0.925	0.164	0.182
	KNEA	0.405	0.927	0.151	0.155
	MARS	0.406	0.925	0.151	0.139
	ANN	0.422	0.908	0.159	0.133
	RF	0.423	0.925	0.159	0.196
	SVM	0.411	0.925	0.155	0.161
	XGBoost	0.425	0.925	0.16	0.184
	HS	0.5	0.854	0.187	0.12
All	ELM	0.338	0.935	0.134	0.043
	GBDT	0.309	0.95	0.122	0.065
	KNEA	0.303	0.947	0.118	0.056
	MARS	0.298	0.948	0.116	0.02
	ANN	0.361	0.934	0.21	− 0.052
	RF	0.328	0.937	0.13	0.043
	SVM	0.3	0.949	0.118	0.061
	XGBoost	0.314	0.944	0.124	0.044
	HS	0.45	0.867	0.177	0.073

Note: the top three ranked models were highlighted in red, green and blue, respectively. Please refer to the online version of this paper to see this table in colour: https://dx.doi.org/10.2166/nh.2019.060.

from the classification tree, but the largest difference between the model and the decision tree is that the base function can be coupled, which has the ability to describe the interaction. This may be the reason why the MARS model is more adaptable. On the other hand, the SVM and KNEA models adopt structural risk minimization and some noise can be artificially ignored by means of tuning parameters, which may explain the high stability of the SVM and KNEA models. Overall, the selection of suitable alternative sites for ET_0 estimation is a systematic project, not only depending on the distance of two sites, but also requiring the similarity of climate rather than the proximity of some individual values. In this study, only four groups of stations (eight stations) were selected to demonstrate the feasibility of switching stations for monthly mean daily ET_0 estimation. However, how to establish a more suitable model still needs to be further explored. In addition, only temperature data were switched in this study and the applicability of using more meteorological data from nearby stations for estimating monthly mean daily ET_0 in a target station is to be studied. Further study is also needed to assess the capability of the proposed models on various time scales (hourly or daily) or in different climatic zones.

CONCLUSIONS

This study compared the capability of eight machine learning models, i.e. ELM, GBDT, KNEA, MARS, ANN, RF, SVM and XGBoost, in modeling monthly mean daily ET_0

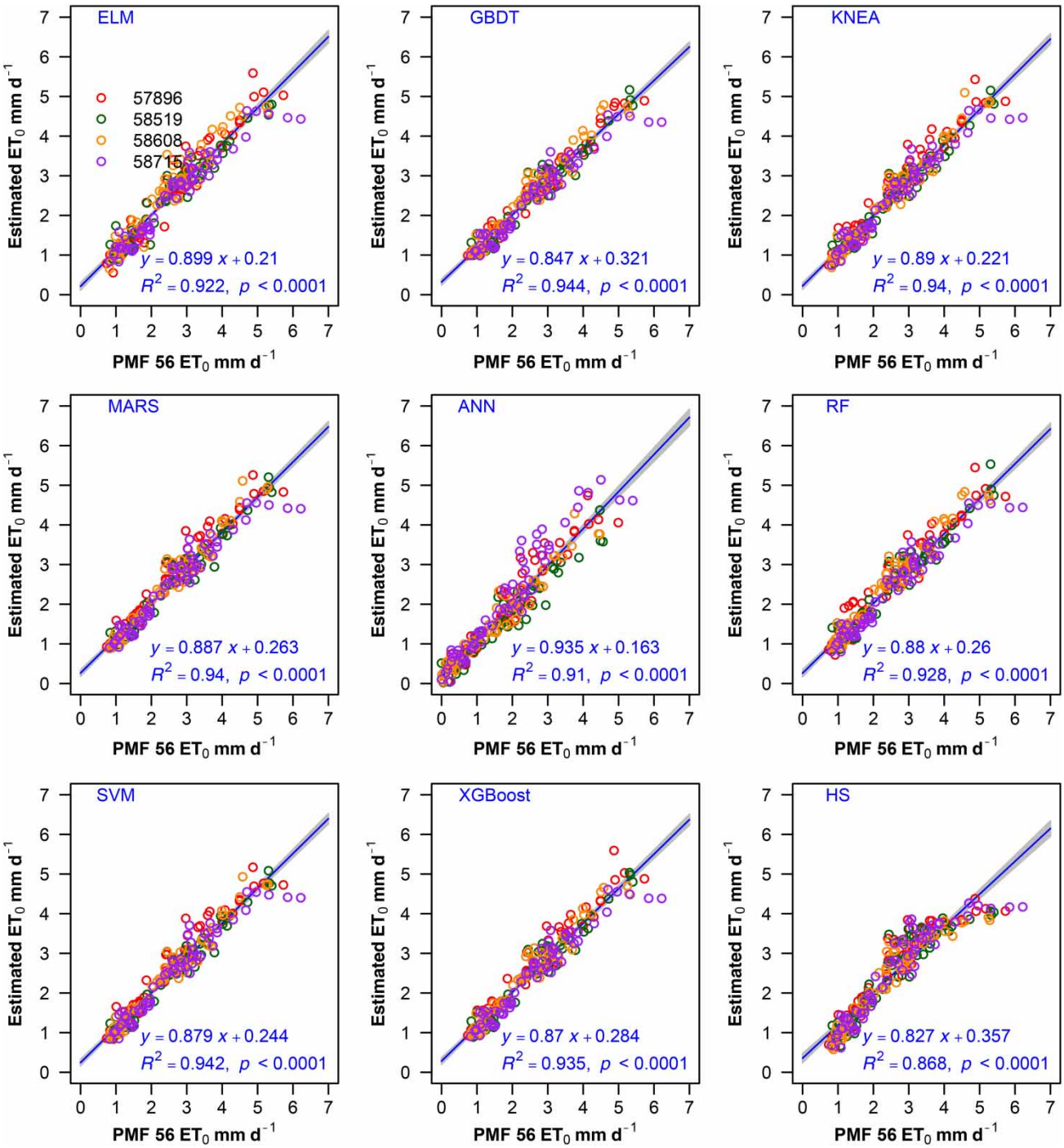

Figure 7 | Scatter plots of the PMF 56 ET_0 and those estimated by the same predictive machine learning models and the HS empirical model performed at the other four stations in cross-station applications.

using maximum and minimum air temperatures and extra-terrestrial solar radiation data from 15 stations located in the Jiangxi Province of China. These machine learning models were also compared with the empirical Hargreaves–Samani model. The results showed that the tree-based models (RF, GBDT and XGBoost) had higher

estimation accuracy than the other models in local applications. When only temperature data were available, the MARS and SVM models performed slightly better than the other models, while the ANN and HS models performed worse than the others. When there was no temperature data at the target station and the temperature data from the adjacent station was used instead, the MARS, SVM and KNEA model outperformed the other models. This study can provide a solution for the estimation of ET_0 in the Jiangxi Province of China when lack of complete meteorological data and may provide a reference for other regions around the world with similar meteorological conditions.

ACKNOWLEDGEMENTS

This study was supported by the National Natural Science Foundation of China (Nos 51879196, 51790533 and 51709143). Thanks to the National Meteorological Information Center of China Meteorological Administration for offering the meteorological data.

REFERENCES

Ali Ghorbani, M., Kazempour, R., Chau, K. W., Shamshirband, S. & Ghazvinei, T. 2018 Forecasting pan evaporation with an integrated artificial neural network quantum-behaved particle swarm optimization model: a case study in Talesh, Northern Iran. *Engineering Applications of Computational Fluid Mechanics* **12** (1), 724–737.

Allen, R. G., Pereira, L. S., Raes, D. & Smith, M. 1998 Crop evapotranspiration – Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56. FAO, Rome, 300(9), D05109.

Almorox, J. & Grieser, J. 2016 Calibration of the Hargreaves–Samani method for the calculation of reference evapotranspiration in different Köppen climate classes. *Hydrology Research* **47** (2), 521–531.

Almorox, J., Quej, V. H. & Martí, P. 2015 Global performance ranking of temperature-based approaches for evapotranspiration estimation considering Köppen climate classes. *Journal of Hydrology* **528**, 514–522.

Beguería, S., Vicente-Serrano, S. M., Reig, F. & Latorre, B. 2014 Standardized precipitation evapotranspiration index (SPEI) revisited: parameter fitting, evapotranspiration models, tools, datasets and drought monitoring. *International Journal of Climatology* **34** (10), 3001–3023.

Breiman, L. 2001 Random forests. *Machine Learning* **45** (1), 5–32.

Chen, T. & Guestrin, C. 2016 Xgboost: a scalable tree boosting system. In: *Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining*. ACM, San Francisco, pp. 785–794.

Cobaner, M., Citakoğlu, H., Haktanir, T. & Kisi, O. 2017 Modifying Hargreaves–Samani equation with meteorological variables for estimation of reference evapotranspiration in Turkey. *Hydrology Research* **48** (2), 480–497.

Ding, C., Wang, D., Ma, X. & Li, H. 2016 Predicting short-term subway ridership and prioritizing its influential factors using gradient boosting decision trees. *Sustainability* **8** (11), 1100.

Elith, J., Leathwick, J. R. & Hastie, T. 2008 A working guide to boosted regression trees. *Journal of Animal Ecology* **77** (4), 802–813.

Fan, J., Wu, L., Zhang, F., Xiang, Y. & Zheng, J. 2016 Climate change effects on reference crop evapotranspiration across different climatic zones of China during 1956–2015. *Journal of Hydrology* **542**, 923–937.

Fan, J., Wang, X., Wu, L., Zhou, H., Zhang, F., Yu, X. & Xiang, Y. 2018a Comparison of support vector machine and extreme gradient boosting for predicting daily global solar radiation using temperature and precipitation in humid subtropical climates: a case study in China. *Energy Conversion and Management* **164**, 102–111.

Fan, J., Chen, B., Wu, L., Zhang, F., Lu, X. & Xiang, Y. 2018b Evaluation and development of temperature-based empirical models for estimating daily global solar radiation in humid regions. *Energy* **144**, 903–914.

Fan, J., Yue, W., Wu, L., Zhang, F., Cai, H., Wang, X. & Xiang, Y. 2018c Evaluation of SVM, ELM and four tree-based ensemble models for predicting daily reference evapotranspiration using limited meteorological data in different climates of China. *Agricultural and Forest Meteorology* **263**, 225–241.

Fan, J., Wu, L., Zhang, F., Cai, H., Ma, X. & Bai, H. 2019a Evaluation and development of empirical models for estimating daily and monthly mean daily diffuse horizontal solar radiation for different climatic regions of China. *Renewable and Sustainable Energy Reviews* **105**, 168–186.

Fan, J., Wu, L., Zhang, F., Cai, H., Zeng, W., Wang, X. & Zou, H. 2019b Empirical and machine learning models for predicting daily global solar radiation from sunshine duration: a review and case study in China. *Renewable and Sustainable Energy Reviews* **100**, 186–212.

Feng, Y., Cui, N., Zhao, L., Hu, X. & Gong, D. 2016 Comparison of ELM, GANN, WNN and empirical models for estimating reference evapotranspiration in humid region of Southwest China. *Journal of Hydrology* **536**, 376–383.

Feng, Y., Jia, Y., Cui, N., Zhao, L., Li, C. & Gong, D. 2017 Calibration of Hargreaves model for reference evapotranspiration estimation in Sichuan basin of southwest China. *Agricultural Water Management* **181**, 1–9.

Feng, Y., Jia, Y., Zhang, Q., Gong, D. & Cui, N. 2018 National-scale assessment of pan evaporation models across different climatic zones of China. *Journal of Hydrology* **564**, 314–328.

Feng, Y., Cui, N., Chen, Y., Gong, D. & Hu, X. 2019 Development of data-driven models for prediction of daily global horizontal irradiance in northwest China. *Journal of Cleaner Production* **223**, 136–146.

Friedman, J. H. 1991 Multivariate adaptive regression splines. *The Annals of Statistics* **19** (1), 1–67.

Gavilán, P., Lorite, I. J., Tornero, S. & Berengena, J. 2006 Regional calibration of Hargreaves equation for estimating reference et in a semiarid environment. *Agricultural Water Management* **81** (3), 0–281.

Ghorbani, M. A., Deo, R. C., Karimi, V., Yaseen, Z. M. & Terzi, O. 2018a Implementation of a hybrid MLP-FFA model for water level prediction of Lake Egirdir, Turkey. *Stochastic Environmental Research and Risk Assessment* **32** (6), 1683–1697.

Ghorbani, M. A., Deo, R. C., Yaseen, Z. M., Kashani, M. H. & Mohammadi, B. 2018b Pan evaporation prediction using a hybrid multilayer perceptron-firefly algorithm (MLP-FFA) model: case study in North Iran. *Theoretical and Applied Climatology* **133** (3–4), 1119–1131.

Hargreaves, G. H. & Allen, R. G. 2003 History and evaluation of Hargreaves evapotranspiration equation. *Journal of Irrigation and Drainage Engineering* **129** (1), 53–63.

Hargreaves, G. H. & Samani, Z. A. 1985 Reference crop evapotranspiration from temperature. *Applied Engineering in Agriculture* **1** (2), 96–99.

Heydari, M. M. & Heydari, M. 2014 Calibration of Hargreaves–Samani equation for estimating reference evapotranspiration in semiarid and arid regions. *Archives of Agronomy and Soil Science* **60** (5), 695–713.

Heydari, M. M., Tajamoli, A., Ghoreishi, S. H., Darbe-Esfahani, M. K. & Gilasi, H. 2015 Evaluation and calibration of Blaney–Criddle equation for estimating reference evapotranspiration in semiarid and arid regions. *Environmental Earth Sciences* **74** (5), 4053–4063.

Hsu, H. M. & Chen, C. T. 1996 Aggregation of fuzzy opinions under group decision making. *Fuzzy Sets and Systems* **79** (3), 279–285.

Huang, G. B., Zhu, Q. Y. & Siew, C. K. 2006 Extreme learning machine: theory and applications. *Neurocomputing* **70** (1–3), 489–501.

Huang, G., Wu, L., Ma, X., Zhang, W., Fan, J., Yu, X. & Zhou, H. 2019 Evaluation of CatBoost method for prediction of reference evapotranspiration in humid regions. *Journal of Hydrology* **574**, 1029–1041.

Huffman, G. J. 1997 Estimates of root-mean-square random error for finite samples of estimated precipitation. *Journal of Applied Meteorology* **36** (9), 1191–1201.

Karimi, S., Kisi, O., Kim, S., Nazemi, A. H. & Shiri, J. 2017 Modelling daily reference evapotranspiration in humid locations of South Korea using local and cross-station data management scenarios. *International Journal of Climatology* **37** (7), 3238–3246.

Khosravi, K., Mao, L., Kisi, O., Yaseen, Z. M. & Shahid, S. 2018 Quantifying hourly suspended sediment load using data mining models: case study of a glacierized Andean catchment in Chile. *Journal of Hydrology* **567**, 165–179.

Kisi, O. 2016 Modeling reference evapotranspiration using three different heuristic regression approaches. *Agricultural Water Management* **169**, 162–172.

Kisi, O. & Alizamir, M. 2018 Modelling reference evapotranspiration using a new wavelet conjunction heuristic method: wavelet extreme learning machine vs wavelet neural networks. *Agricultural and Forest Meteorology* **263**, 41–48.

Landeras, G., Bekoe, E., Ampofo, J., Logah, F., Diop, M., Cisse, M. & Shiri, J. 2018 New alternatives for reference evapotranspiration estimation in West Africa using limited weather data and ancillary data supply strategies. *Theoretical and Applied Climatology* **132** (3–4), 701–716.

Liu, W., Yang, H., Folberth, C., Wang, X., Luo, Q. & Schulin, R. 2016 Global investigation of impacts of PET methods on simulating crop-water relations for maize. *Agricultural and Forest Meteorology* **221**, 164–175.

Lu, X., Ju, Y., Wu, L., Fan, J., Zhang, F. & Li, Z. 2018 Daily pan evaporation modeling from local and cross-station data using three tree-based machine learning models. *Journal of Hydrology* **566**, 668–684.

Luo, Y., Chang, X., Peng, S., Khan, S., Wang, W., Zheng, Q. & Cai, X. 2014 Short-term forecasting of daily reference evapotranspiration using the Hargreaves–Samani model and temperature forecasts. *Agricultural Water Management* **136**, 42–51.

Luo, Y., Traore, S., Lyu, X., Wang, W., Wang, Y., Xie, Y., Jiao, X. & Fipps, G. 2015 Medium range daily reference evapotranspiration forecasting by using ANN and public weather forecasts. *Water Resources Management* **29** (10), 3863–3876.

Ma, X. 2019 A brief introduction to the Grey Machine Learning. *Journal of Grey Systems* **31** (1), 1–12.

Ma, X. & Liu, Z. 2018a Predicting the oil production using the novel multivariate nonlinear model based on Arps decline model and kernel method. *Neural Computing and Applications* **29** (2), 579–591.

Ma, X. & Liu, Z. B. 2018b The kernel-based nonlinear multivariate grey model. *Applied Mathematical Modelling* **56**, 217–238.

Ma, X., Xie, M., Wu, W., Zeng, B., Wang, Y. & Wu, X. 2019a The novel fractional discrete multivariate grey system model and its applications. *Applied Mathematical Modelling* **70**, 402–424.

Ma, X., Mei, X., Wu, W., Wu, X. & Zeng, B. 2019b A novel fractional time delayed grey model with Grey Wolf Optimizer and its applications in forecasting the natural gas and coal consumption in Chongqing China. *Energy* **178**, 487–507.

Martel, M., Glenn, A., Wilson, H. & Kröbel, R. 2018 Simulation of actual evapotranspiration from agricultural landscapes in the Canadian Prairies. *Journal of Hydrology: Regional Studies* **15**, 105–118.

Martí, P. & Gasque, M. 2011 Improvement of temperature-based ANN models for solar radiation estimation through exogenous data assistance. *Energy Conversion and Management* **52** (2), 990–1003.

Martí, P. & Zarzo, M. 2012 Multivariate statistical monitoring of ETo: A new approach for estimation in nearby locations using geographical inputs. *Agricultural and Forest Meteorology* **152**, 125–134.

Martí, P., Zarzo, M., Vanderlinden, K. & Girona, J. 2015 Parametric expressions for the adjusted Hargreaves coefficient in Eastern Spain. *Journal of Hydrology* **529**, 1713–1724.

Mattar, M. A. 2018 Using gene expression programming in monthly reference evapotranspiration modeling: a case study in Egypt. *Agricultural Water Management* **198**, 28–38.

Mccabe, G. J., Hay, L. E., Bock, A., Markstrom, S. L. & Atkinson, R. D. 2015 Inter-annual and spatial variability of hamon potential evapotranspiration model coefficients. *Journal of Hydrology* **521**, 389–394.

Mehdizadeh, S., Behmanesh, J. & Khalili, K. 2017 Using MARS, SVM, GEP and empirical equations for estimation of monthly mean reference evapotranspiration. *Computers and Electronics in Agriculture* **139**, 103–114.

Moazenzadeh, R., Mohammadi, B., Shamshirband, S. & Chau, K. W. 2018 Coupling a firefly algorithm with support vector regression to predict evaporation in northern Iran. *Engineering Applications of Computational Fluid Mechanics* **12** (1), 584–597.

Mohan, A., Chen, Z. & Weinberger, K. 2011 Web-search ranking with initialized gradient boosted regression trees. In: *Proceedings of the Learning to Rank Challenge*, pp. 77–89.

Morales-Salinas, L., Ortega-Farías, S., Riveros-Burgos, C., Neira-Román, J., Carrasco-Benavides, M. & López-Olivari, R. 2017 Monthly calibration of Hargreaves–Samani equation using remote sensing and topoclimatology in central-southern Chile. *International Journal of Remote Sensing* **38** (24), 7497–7513.

Naganna, S. R., Deka, P. C., Ghorbani, M. A., Biazar, S. M., Al-Ansari, N. & Yaseen, Z. M. 2019 Dew point temperature estimation: application of artificial intelligence model integrated with nature-inspired optimization algorithms. *Water* **11** (4), 742.

Oudin, L., Michel, C. & Anctil, F. 2005 Which potential evapotranspiration input for a lumped rainfall-runoff model?: Part 1—can rainfall-runoff models effectively handle detailed potential evapotranspiration inputs? *Journal of Hydrology* **303** (1–4), 275–289.

Pandey, V., Pandey, P. K. & Mahanta, A. P. 2014 Calibration and performance verification of Hargreaves Samani equation in a humid region. *Irrigation and Drainage* **63** (5), 659–667.

Pereira, A. R. & Pruitt, W. O. 2004 Adaptation of the Thornthwaite scheme for estimating daily reference evapotranspiration. *Agricultural Water Management* **66** (3), 251–257.

Quej, V. H., Almorox, J., Arnaldo, J. A. & Saito, L. 2017 ANFIS, SVM and ANN soft-computing techniques to estimate daily global solar radiation in a warm sub-humid environment. *Journal of Atmospheric and Solar-Terrestrial Physics* **155**, 62–70.

Quej, V. H., Almorox, J., Arnaldo, J. A. & Moratiel, R. 2018 Evaluation of temperature-based methods for the estimation of reference evapotranspiration in the Yucatán peninsula, Mexico. *Journal of Hydrologic Engineering* **24** (2), 05018029.

Ravazzani, G., Corbari, C., Morella, S., Gianoli, P. & Mancini, M. 2011 Modified Hargreaves-Samani equation for the assessment of reference evapotranspiration in Alpine river basins. *Journal of Irrigation and Drainage Engineering* **138** (7), 592–599.

Raziei, T. & Pereira, L. S. 2013 Estimation of ETo with Hargreaves–Samani and FAO-PM temperature methods for a wide range of climates in Iran. *Agricultural Water Management* **121**, 1–18.

Samani, Z. 2000 Estimating solar radiation and evapotranspiration using minimum climatological data. *Journal of Irrigation and Drainage Engineering* **126** (4), 265–267.

Sanikhani, H., Kisi, O., Maroufpoor, E. & Yaseen, Z. M. 2019 Temperature-based modeling of reference evapotranspiration using several artificial intelligence models: application of different modeling scenarios. *Theoretical and Applied Climatology* **135** (1–2), 449–462.

Seiller, G. & Anctil, F. 2016 How do potential evapotranspiration formulas influence hydrological projections? *Hydrological Sciences Journal* **61** (12), 2249–2266.

Shahidian, S., Serralheiro, R. P., Serrano, J. & Teixeira, J. L. 2013 Parametric calibration of the Hargreaves–Samani equation for use at new locations. *Hydrological Processes* **27** (4), 605–616.

Shiri, J. 2017 Evaluation of FAO56-PM, empirical, semi-empirical and gene expression programming approaches for estimating daily reference evapotranspiration in hyper-arid regions of Iran. *Agricultural Water Management* **188**, 101–114.

Shiri, J. 2018 Improving the performance of the mass transfer-based reference evapotranspiration estimation approaches through a coupled wavelet-random forest methodology. *Journal of Hydrology* **561**, 737–750.

Shiri, J. 2019 Modeling reference evapotranspiration in island environments: assessing the practical implications. *Journal of Hydrology* **570**, 265–280.

Shiri, J., Nazemi, A. H., Sadraddini, A. A., Landeras, G., Kisi, O., Fard, A. F. & Marti, P. 2014 Comparison of heuristic and empirical approaches for estimating reference evapotranspiration from limited inputs in Iran. *Computers and Electronics in Agriculture* **108**, 230–241.

Shiri, J., Marti, P., Nazemi, A. H., Sadraddini, A. A., Kisi, O., Landeras, G. & Fakheri Fard, A. 2015 Local vs. external training of neuro-fuzzy and neural networks models for estimating reference evapotranspiration assessed through k-fold testing. *Hydrology Research* **46** (1), 72–88.

Shiri, J., Marti, P., Karimi, S. & Landeras, G. 2019 Data splitting strategies for improving data driven models for reference evapotranspiration estimation among similar stations. *Computers and Electronics in Agriculture* **163**, 70–81.

Tao, H., Diop, L., Bodian, A., Djaman, K., Ndiaye, P. M. & Yaseen, Z. M. 2018a Reference evapotranspiration prediction using hybridized fuzzy model with firefly algorithm: regional case study in Burkina Faso. *Agricultural Water Management* **208**, 140–151.

Tao, H., Sulaiman, S. O., Yaseen, Z. M., Asadi, H., Meshram, S. G. & Ghorbani, M. A. 2018b What is the potential of integrating phase space reconstruction with SVM-FFA data-intelligence model? Application of rainfall forecasting over regional scale. *Water Resources Management* **32** (12), 3935–3959.

Trajkovic, S. 2005 Temperature-based approaches for estimating reference evapotranspiration. *Journal of Irrigation and Drainage Engineering* **131** (4), 316–323.

Valiantzas, J. D. 2017 Modification of the Hargreaves–Samani model for estimating solar radiation from temperature and humidity data. *Journal of Irrigation and Drainage Engineering* **144** (1), 06017014.

Valipour, M. 2015 Importance of solar radiation, temperature, relative humidity, and wind speed for calculation of reference evapotranspiration. *Archives of Agronomy and Soil Science* **61** (2), 239–255.

Valipour, M., Sefidkouhi, M. A. G. & Raeini, M. 2017 Selecting the best model to estimate potential evapotranspiration with respect to climate change and magnitudes of extreme events. *Agricultural Water Management* **180**, 50–60.

Vapnik, V. 2013 *The Nature of Statistical Learning Theory*. Springer Science & Business Media, Berlin.

Wu, L., Huang, G., Fan, J., Zhang, F., Wang, X. & Zeng, W. 2019 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 Conversion and Management* **183**, 280–295.

Xiao, Q., Li, C., Tang, Y., Li, L. & Li, L. 2019 A knowledge-driven method of adaptively optimizing process parameters for energy efficient turning. *Energy* **166**, 142–156.

Xu, J., Wang, J., Wei, Q. & Wang, Y. 2016 Symbolic regression equations for calculating daily reference evapotranspiration with the same input to Hargreaves-Samani in arid China. *Water Resources Management* **30** (6), 2055–2073.

Yaseen, Z. M., Allawi, M. F., Yousif, A. A., Jaafar, O., Hamzah, F. M. & El-Shafie, A. 2018 Non-tuned machine learning approach for hydrological time series forecasting. *Neural Computing and Applications* **30** (5), 1479–1491.

Yaseen, Z. M., Fu, M., Wang, C., Mohtar, W. H. M. W., Deo, R. C. & El-Shafie, A. 2018a Application of the hybrid artificial neural network coupled with rolling mechanism and grey model algorithms for streamflow forecasting over multiple time horizons. *Water Resources Management* **32** (5), 1883–1899.

Yaseen, Z. M., Ghareb, M. I., Ebtehaj, I., Bonakdari, H., Siddique, R., Heddam, S. & Deo, R. 2018b Rainfall pattern forecasting using novel hybrid intelligent model based ANFIS-FFA. *Water Resources Management* **32** (1), 105–122.

Yaseen, Z. M., Sulaiman, S. O., Deo, R. C. & Chau, K. W. 2018c An enhanced extreme learning machine model for river flow forecasting: State-of-the-art, practical applications in water resource engineering area and future research direction. *Journal of Hydrology* **569**, 387–408.

Yassin, M. A., Alazba, A. A. & Mattar, M. A. 2016 Artificial neural networks versus gene expression programming for estimating reference evapotranspiration in arid climate. *Agricultural Water Management* **163**, 110–124.

Zhao, L., Xia, J., Xu, C. Y., Wang, Z., Sobkowiak, L. & Long, C. 2013 Evapotranspiration estimation methods in hydrological models. *Journal of Geographical Sciences* **23** (2), 359–369.

First received 4 May 2019; accepted in revised form 12 September 2019. Available online 1 October 2019