https://doi.org/10.5194/hess-2023-158
Preprint. Discussion started: 31 July 2023
© Author(s) 2023. CC BY 4.0 License.

Hydrology and Earth System Sciences
Discussions
Open Access

# Comparison of four machine learning models for forecasting daily reference evaporation based on public weather forecast data

Yunfeng Liang [a,b] , Dongpu Feng [a,c] , Zhaojun Sun [a, d, e, f*]

a School of Civil and Hydraulic Engineering, Ningxia University, Yinchuan 750021, Ningxia, China

b School of Mechanical Engineering, Ningxia University, Yinchuan 750021, Ningxia, China

c Engineering Research Center for Efficient Utilization of Modern Agricultural Water Resources in Arid Regions, Ministry of Education, Ningxia University, Yinchuan, 750021, Ningxia, China

d School of Geography and Planning, Ningxia University, Yinchuan 750021, Ningxia, China

e China-Arab Joint International Research Laboratory for Featured Resources and Environmental Governance in Arid Region, Yinchuan 750021, Ningxia, China

f Key Laboratory of Resource Assessment and Environmental Control in Arid Region of Ningxia, Yinchuan 750021, Ningxia, China

* Corresponding author at: School of Geography and Planning, Ningxia University, Yinchuan 750021, Ningxia, China.

E-mail address: zhaojunsun_nx@163.com (Z. Sun).

https://doi.org/10.5194/hess-2023-158
Preprint. Discussion started: 31 July 2023
© Author(s) 2023. CC BY 4.0 License.



Abstract: Real-time accurate prediction of daily reference evapotranspiration ($ET_o$) is critical for real-time irrigation decisions and water resource management. Although many public weather forecast-based machine learning models have been successfully used for daily $ET_o$ prediction, these models are developed with long-term historical daily observed meteorological data. The use of training and testing samples from different data sources can lead to the selection of the best model, and the performance of the best model for predicting daily $ET_o$ is not ideal. In this study, based on Food and Agriculture Organization (FAO) 56 Penman–Monteith (PM) equations, four machine learning models (multilayer perceptron ($MLP_o$), extreme gradient boosting ($XGBoost_o$), light gradient boosting machine ($LightGBM_o$), and gradient boosting with categorical features support ($CatBoost1_o$)) were trained and validated with daily observed meteorological data from 1995-2015 and 2016-2019, respectively, and five machine learning models ($MLP_p$, $XGBoost_p$, $LightGBM_p$, $CatBoost1_p$, and CatBoost2) were trained and validated with daily public weather forecast data with a 1-day lead time (2014-2018 and 2019, respectively). Based on public weather forecast and daily observed meteorological data (2020-2021), the predicted daily $ET_o$ performance of nine machine learning models ($MLP_o$, $XGBoost_o$, $LightGBM_o$, $CatBoost1_o$, $MLP_p$, $XGBoost_p$, $LightGBM_p$, $CatBoost1_p$, and CatBoost2) was compared. The results show that for all three studied climate zones, the performance of the four models developed based on public weather forecast data with a 1-day advance is better than that of the four models developed based on daily observed meteorological data with corresponding input combinations, and the mean MAE and RMSE ranges for the four models (MLP, XGBoost, LightGBM, and CatBoost1) in the three studied climate zones were reduced by 2.93%-11.67% and 2.20%-9.46%, respectively, and the mean R range was improved by 1.31%-5.31%. The top three models for the AR climate zone were $XGBoost_p$, $LightGBM_p$, and

https://doi.org/10.5194/hess-2023-158
Preprint. Discussion started: 31 July 2023
© Author(s) 2023. CC BY 4.0 License.



$MLP_p$, the top three models for the SAR climate zone were $MLP_p$, $XGBoost_p$, and $LightGBM_p$, and the top three models for the SHZ climate zone were $XGBoost_p$, $MLP_p$, and $LightGBM_p$. In addition, the prediction performance for daily $ET_o$ is found to be highest in winter and lowest in summer in all three climate zones. *Wspd* from public weather forecasts was the most important source of daily $ET_o$ error in model predictions for the AR climate zone, followed by *SDun*, $T_{max}$, and $T_{min}$, while *SDun* from public weather forecasts was the most important source of daily $ET_o$ error in model predictions for the SAR (SHZ) climate zone, followed by *Wspd*, $T_{max}$, and $T_{min}$ ($T_{max}$, *Wspd*, and $T_{min}$).

Keywords: Forecasting, Reference evapotranspiration, Public weather forecast, Tree-based assembly algorithms, Irrigation season

## 1. Introduction

Crop water demand is the most important part of water transfer and energy conversion in the soil–plant–atmosphere continuum (SPAC), and it is an important process in the field water cycle. Accurately estimating crop water demand is the basis for designing crop irrigation systems, determining regional irrigation water use, and facilitating effective basin planning, regional water planning, and drainage and irrigation project planning, design, and management. However, directly measuring crop water demand is time-consuming and expensive, significantly limiting practical applications (Irmak et al., 2003; Martınez-Cob et al., 2015; Silva et al. 2019a; Fan et al., 2021a). Thus, crop water demand is usually measured using an indirect method, i.e., reference evapotranspiration ($ET_o$) multiplied by the crop coefficient. Predicting $ET_o$ is more valuable than estimating $ET_o$ (Yang et al., 2016; Yang et al., 2019a), and accurate $ET_o$ prediction is the key to crop

https://doi.org/10.5194/hess-2023-158
Preprint. Discussion started: 31 July 2023
© Author(s) 2023. CC BY 4.0 License.



water demand prediction and the prerequisite for real-time irrigation forecasting, which has important reference value and significance for real-time irrigation decisions (Luo et al., 2014; Perera et al. 2014; Ballesteros et al., 2016).

Depending on the method and input data, $ET_o$ prediction methods can be divided into direct and indirect methods (Perera et al., 2014). Since daily $ET_o$ data mainly vary with weather and are only influenced by future weather variables, direct methods that use long-term historical weather data to predict $ET_o$ are not applicable to short-term daily $ET_o$ prediction. For indirect $ET_o$ prediction methods that use weather forecast data, numerical weather forecasts that provide all the variables of the Food and Agriculture Organization 56 Penman–Monteith (FAO 56 PM) equation have been used in predicting daily $ET_o$ (Silva et al., 2010; Pelosi et al., 2016), and used in predicting daily $ET_o$ with a 1-10 day lead time (Perera et al. 2014; Medina et al., 2018; Vanella et al., 2020). However, numerical forecast products in China are only available to professionals (e.g., registered users in academia and research) and are not accessible to nonprofessionals. In addition, these products require preprocessing (Fan et al., 2021b) or postprocessing (Medina et al., 2020) by professionals to improve the reliability of the output data.

In China, the China Weather Network, hosted by the Public Meteorological Service Center of the China Meteorological Administration, is a public service-based meteorological portal for society and the public. The China Weather Network covers more than 100,000 sites in provinces, cities, towns, and tourist attractions across the country and provides real-time meteorological services, including weather forecasts, current conditions, indices, air quality information, and many other elements, with a minimum time resolution of 5 minutes and a maximum forecast time of 40 days. It also provides user location-based forecasting services on mobile sites. Public weather forecasts

https://doi.org/10.5194/hess-2023-158
Preprint. Discussion started: 31 July 2023
© Author(s) 2023. CC BY 4.0 License.


include four variables: maximum temperature, minimum temperature, wind scale and weather type. In recent years, public weather forecasts have been widely used in $ET_o$ forecasting. For example, information from public weather forecasts has been converted into variables required to calculate ETo using the FAO-56 PM equation, and then used to forecast in daily $ET_o$ (Cai et al., 2007; Cai et al., 2009), used to forecast daily $ET_o$ with a lead time of 1-3 days (Liu et al., 2020), and used to forecast daily $ET_o$ with a lead time of 1-7 days (Yang et al., 2016). The results of these studies indicate that the errors in daily $ET_o$ predicted with the variables transformed with information from public weather forecasts and the FAO 56 PM equation are mainly caused by errors arising from the process of converting qualitative wind scale and weather type information in public weather forecasts into wind speed and sunshine hour information, respectively.

Thus, some studies have included measured or precalculated variable data and public weather forecast data as inputs for their models to predict daily $ET_o$. An example of such studies include using temperature data from weather forecasts and actual incident net solar radiation (Rs) values occurring in each advance period and four artificial neural network (ANN) learning algorithms (generalized feedforward (GFF), linear regression (LR), multilayer perceptron (MLP) and probabilistic neural network (PNN)) to forecast $ET_o$ in Dallas, Texas, USA, 1-15 days in advance (Traore et al., 2016). Additionally, temperature data from public weather forecasts, measured sunshine hours, precalculated weather type correction factors (Wt, only four weather type correction factors, i.e., sunny, cloudy, overcast and rainy, are defined) and a combined model of bi-directional long short-term memory (Bi-LSTM) and ANN were used to predict daily $ET_o$ 1-7 days ahead for three stations in Ningxia, China (Yin et al., 2020). However, since short-term daily $ET_o$ is mainly governed by weather conditions, introducing additional measured or precalculated variable data as

https://doi.org/10.5194/hess-2023-158
Preprint. Discussion started: 31 July 2023
© Author(s) 2023. CC BY 4.0 License.



input to predict daily $ET_o$ may not be applicable for real-time irrigation decisions. In addition, in these studies, machine learning models were trained and validated with long-term historical meteorological data and tested with public weather forecast data (with a maximum data duration of 2 years). Thus, the data sources in these studies were different (e.g., in China, the source of historical meteorological data is the National Meteorological Information Center of China, while the source of public weather forecasts is the Public Weather Service Center of the China Meteorological Administration). Moreover, there have been no comparison studies between training and validating machine learning models with historical meteorological data and public weather forecast data. However, the use of samples from different data sources may affect the daily $ET_o$ performance of machine learning models.

In recent years, three integrated learning models, extreme gradient boosting (XGBoost), light gradient boosting machine (LightGBM) and gradient boosting with categorical features support (CatBoost), based on a boosting algorithm with a decision tree as the base learner and greedy ideas for decision tree growth, have been widely used to estimate daily $ET_o$. Among them, CatBoost can directly handle categorical data, such as public weather forecasts of wind scale and weather type. Fan et al. (2018) recommended using the XGBoost and GBDT models with limited climate data to predict daily $ET_o$ under different climatic conditions in China. Fan et al. (2019) showed that LightGBM generally outperformed the other two soft computing models and four empirical models in estimating daily $ET_o$ with data from 49 weather stations in humid subtropical China using local meteorological data and cross-station meteorological data. Zhou et al. (2020) strongly recommended the use of CatBoost and LightGBM models for estimating daily $ET_o$ under different climatic conditions in China. Huang et al. (2019) showed that the CatBoost algorithm has great potential for

https://doi.org/10.5194/hess-2023-158
Preprint. Discussion started: 31 July 2023
© Author(s) 2023. CC BY 4.0 License.


daily $ET_o$ estimation in humid regions of China. Zhang et al. (2020) also showed that CatBoost is considered the best choice for estimating $ET_o$ in arid and semiarid regions of northern China. These studies showed that all three models estimated daily $ET_o$ with better accuracy and stability than those of other models. However, studies on the prediction of daily $ET_o$ with these three models have not been reported. The objectives of this study are as follows: (1) to first develop four machine learning models using daily observed meteorological data and public weather forecast data with a 1-day lead time, respectively, and then to test and compare the performance of the developed models in both cases using public weather forecast data with a 1-7 day lead time; (2) to explore the use of categorical data (public weather forecasts of wind scale and weather type) as direct input into CatBoost for daily $ET_o$ prediction; and (3) to compare the seasonal variation in predicted daily $ET_o$ performance of the five machine learning models and recommend the best daily $ET_o$ prediction model for all four seasons at nine stations in three different climates.

## 2. Materials and methodology

### 2.1. Study area and data collection

Ningxia is located between 35°25′-39°25′ N latitude and 104°10′-107°30′ E longitude and has a temperate continental climate. According to the Köppen classification (Kottek et al., 2006), the climate of Ningxia is divided into three climatic zones, a mid-temperate arid zone in the north (Northern Yellow Irrigation Zone), a mid-temperate semiarid zone in the middle (Central Arid Zone) and a mid-temperate semihumid zone in the south (Southern Mountainous Zone). A total of nine meteorological stations were selected from these three different climatic zones. The geographical distribution of the study sites is shown in Fig. 1, and the characteristics of the study sites are shown in Table 1.

https://doi.org/10.5194/hess-2023-158
Preprint. Discussion started: 31 July 2023
© Author(s) 2023. CC BY 4.0 License.
Hydrology and Earth System Sciences
Discussions
Open Access

**Table 1**

Characteristics of the nine weather stations used in this study.

| No. | Station No. | Station | Climate zone | Latitude (°N) | Longitude (°E) | Elevation (m) | $T_{max}$ (℃) | $T_{min}$ (℃) | SDun (h) | $u_2$ ($m\ s^{-1}$) | $RH_{mean}$ (%) | $ET_o$ ($mm\ d^{-1}$) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 53519 | Huinong(HN) | AR | 39.22 | 106.77 | 1092.2 | 17.63 | 4.78 | 8.27 | 1.63 | 44.15 | 3.23 |
| 2 | 53614 | Yinchuan(YC) | AR | 38.47 | 106.20 | 1111.6 | 17.66 | 5.23 | 7.53 | 1.54 | 48.31 | 2.94 |
| 3 | 53704 | Zhongwei(ZW) | AR | 37.53 | 105.18 | 1226.7 | 18.06 | 4.14 | 7.87 | 1.79 | 51.60 | 3.19 |
| 4 | 53705 | Zhongning(ZN) | AR | 37.48 | 105.68 | 1181.3 | 18.53 | 5.88 | 7.70 | 1.49 | 48.25 | 3.18 |
| 5 | 53723 | Yanchi(YAC) | SAR | 37.80 | 107.23 | 1350.9 | 16.73 | 3.14 | 7.64 | 2.37 | 50.39 | 2.97 |
| 6 | 53806 | Haiyuan(HY) | SAR | 36.57 | 105.65 | 1855.6 | 14.51 | 3.48 | 6.92 | 1.83 | 50.64 | 2.88 |
| 7 | 53810 | Tongxin(TX) | SAR | 36.97 | 105.90 | 1336.4 | 17.89 | 4.96 | 7.71 | 2.28 | 51.57 | 3.45 |
| 8 | 53817 | Guyuan(GY) | SHZ | 36.03 | 106.23 | 1835.5 | 14.16 | 3.26 | 6.65 | 1.87 | 56.96 | 2.68 |
| 9 | 53903 | Xiji(XJ) | SHZ | 35.97 | 105.72 | 1916.5 | 13.82 | 1.08 | 6.01 | 1.17 | 63.73 | 2.33 |

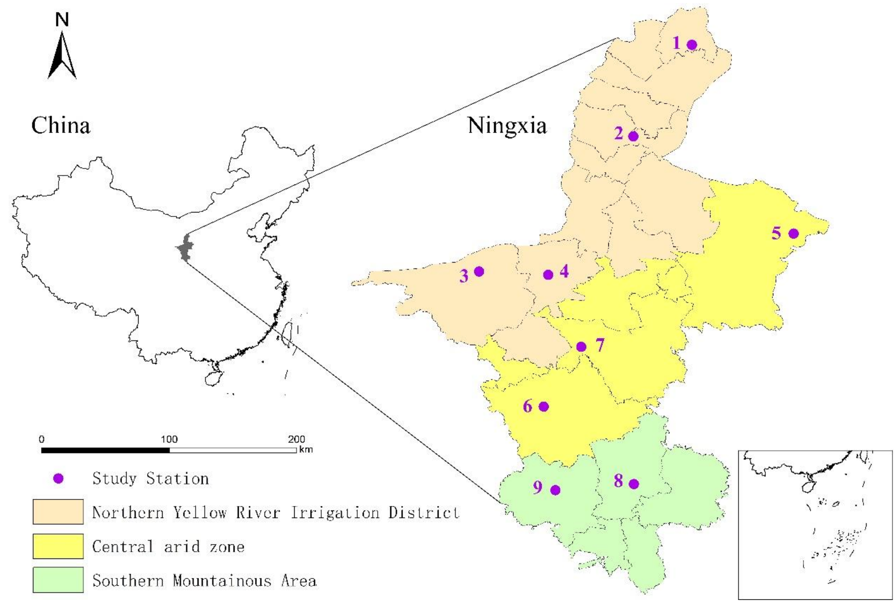



**Fig. 1.** Locations of the weather stations used in this study

The daily observed meteorological data, including the daily maximum temperature ($T_{max}$), daily minimum temperature ($T_{min}$), average temperature, average relative humidity, sunshine hours (SDun), and average wind speed (Wspd), were obtained from the China Meteorological Data Network (http://data.cma.cn/) for the nine meteorological stations during the period from January 1, 1995, to December 31, 2021. Public weather forecast data during 1-7 days ahead from January 1, 2014, to December 31, 2021, including the daily maximum temperature ($T_{max}$), daily minimum

https://doi.org/10.5194/hess-2023-158
Preprint. Discussion started: 31 July 2023
© Author(s) 2023. CC BY 4.0 License.


temperature ($T_{min}$), wind scale (WS), and weather type (WT), were collected from the China Weather Network (http://www.weather.com.cn) for the same stations.

Daily observed weather data from 1995-2015 and 2016-2019 were used for the training and validation of four machine learning models, respectively. Public weather forecast data with 1-day lead times from 2014-2018 and 2019 were used for the training and validation of five machine learning models, respectively. The performance of the 9 developed machine learning models was tested with public weather forecast data with a 1-7 day lead time from 2020-2021 and compared with the daily $ET_o$ calculated by the FAO-56 PM equation and daily observed meteorological data from 2020-2021.

## 2.2 Methodology

### 2.2.1 Food and Agriculture Organization 56 Penman–Monteith equation

The FAO 56 PM equation (Allen et al., 1998), recommended by the United Nations Food and Agriculture Organization, is given as follows to assess the performance of machine learning models in terms of predicting $ET_o$.

$$\mathrm{ET_o} = \frac{0.408\Delta(R_n - G) + \gamma \frac{900}{T + 273} u_2 (e_s - e_a)}{\Delta + \gamma(1 + 0.34u_2)} \quad (1)$$

where $ET_o$ is the daily reference evapotranspiration [mm day $^{-1}$]; $R_n$ is the net radiation at the crop surface [MJ m$^{-2}$ day$^{-1}$]; G is the soil heat flux density [MJ m$^{-2}$ day$^{-1}$], and G may be ignored for day periods; T is the mean daily air temperature at 2 m height [°C]; $u_2$ is the wind speed at 2 m height [m s$^{-1}$]; $e_s$ is the saturation vapour pressure [kPa]; $e_a$ is the actual vapour pressure [kPa]; es -ea is the saturation vapour pressure deficit [kPa]; Δ is the slope vapour pressure curve [kPa °C$^{-1}$]; $\gamma$ is the psychrometric constant [kPa °C$^{-1}$].

### 2.2.2. Machine learning models

https://doi.org/10.5194/hess-2023-158
Preprint. Discussion started: 31 July 2023
© Author(s) 2023. CC BY 4.0 License.
Hydrology and Earth System Sciences
Open Access
Discussions

The two main types of integrated learning algorithms are bagging-based algorithms and boosting-based algorithms. XGBoost, LightGBM and CatBoost are improved implementations of the GBDT algorithm and belong to the boosting algorithm family. Among them, XGBoost was proposed by Chen et al. (2016), and its official open source documentation is available at http://xgboost.readthedocs.io. LightGBM was proposed by Ke et al. (2017), and its official open source documentation is available at http://lightgbm.readthedocs.io. CatBoost was proposed by Prokhorenkova et al. (2017), and its official open source documentation is available at https://catboost.ai/en/docs/. The base learners of all three algorithms are decision trees, but the tree features and the process of generation differ in many ways. For example, XGBoost uses a levelwise decision tree growth strategy, LightGBM uses a leafwise decision tree growth strategy with depth restrictions, and CatBoost uses a fully symmetric decision tree as the base learner.

For categorical features, CatBoost only needs to declare the categorical signs to enable direct feature processing (Prokhorenkova et al. 2017; Dorogush et al. 2018). LightGBM associates each categorical feature fetch with a bucket (bin) and thus automatically processes the features without preprocessing using one-hot encoding. XGBoost cannot process the categorical features directly. This model is used after preprocessing the categorical features by various encoding methods such as tag encoding, mean encoding, or one-hot encoding.

The concept of deep learning was introduced by Hinton et al. (2006). A multilayer perceptron (MLP) with multiple hidden layers is a deep learning structure. MLP with one or two hidden layers has been successfully used for $ET_o$ estimation or prediction. Landeras et al. (2008) used an artificial neural network (ANN) with 1 hidden layer to estimate the daily $ET_o$ for Alava, Basque Country in northern Spain and obtained better results than those from 10 locally calibrated empirical and

https://doi.org/10.5194/hess-2023-158
Preprint. Discussion started: 31 July 2023
© Author(s) 2023. CC BY 4.0 License.



semiempirical $ET_o$ equations and their variants. Ferreira et al. (2019) estimated daily $ET_o$ for all of Brazil using the first four days of data, and an ANN (model structure 16-50-50-1) was the best choice among temperature- and relative humidity-based models. The ANN with two hidden layers used by Elbeltagi et al. (2022) is a suitable alternative to estimate the daily $ET_o$ for the meteorological station in Debrecen, Hungary, based on limited meteorological data. Luo et al. (2015) used an MLP with two hidden layers and public weather forecast data with a 1-7 day lead time to predict the daily $ET_o$ for Gaoyou station in Jiangsu Province, China, with acceptable prediction performance. Traore et al. (2016) showed that an MLP model with two hidden layers and a combination of $T_{max}$, $T_{min}$, and $R_s$ inputs had the best $ET_o$ prediction performance for Dallas, Texas, USA. These studies used trial-and-error methods to determine the numbers of hidden layers and neurons per hidden layer. These methods are time-consuming and may not always yield the best hyperparameters. In this study, an MLP with multiple hidden layers is compared with three machine learning models, XGBoost, LightGBM and CatBoost. The MLP with multiple hidden layers is implemented using Google's TensorFlow deep learning framework. See Figure 2 for details.

### 2.2.3. Input combinations and hyperparameter tuning methods for the machine learning models

The wind scale (WS) data in public weather forecast information are converted to wind speed data (Wspd), as shown in Table 2, and the weather type information is converted using the analytical method proposed by Cai et al. (2007). First, the weather type information is converted to the sunshine hour coefficient, as shown in Table 3, and then the predicted sunshine hour (SDun) is obtained by combining with Equation (3) (Allen et al., 1998, Cai et al., 2007, Yang et al., 2016, Yang et al., 2019b).

$$N = \frac{24}{\pi}\omega_s \quad (2)$$

https://doi.org/10.5194/hess-2023-158
Preprint. Discussion started: 31 July 2023
© Author(s) 2023. CC BY 4.0 License.


$$\mathrm{SDun} = \alpha N \tag{3}$$

where N is the daylight hours [h]; $\omega_s$ is the sunset hour angle [rad]; n is the predicted duration of sunshine [h]; α is the coefficient of sunshine duration.

**Table 2**

Beaufort wind scale (GB/T 35227—2017, 2017).

| Wind scale(**WS**) | Designation | $u_{10}$($ms^{-1}$) | |
|---|---|---|---|
| | | Range | Average(**Wspd**) |
| 0 | Calm | 0.0 – 0.2 | 0.0 |
| 1 | Light | 0.3 – 1.5 | 1.0 |
| 2 | Slight | 1.6 – 3.3 | 2.0 |
| 3 | Gentle | 3.4 – 5.4 | 4.0 |
| 4 | Moderate | 5.5 – 7.9 | 7.0 |
| 5 | Fresh | 8.0 – 10.7 | 9.0 |
| 6 | Strong wind | 10.8 – 13.8 | 12.0 |
| 7 | High wind | 13.9 – 17.1 | 16.0 |
| 8 | Gale | 17.2 – 20.7 | 19.0 |
| 9 | Strong gale | 20.8 – 24.4 | 23.0 |
| 10 | Whole gale | 24.5 – 28.4 | 26.0 |
| 11 | Storm | 28.5 – 32.6 | 31.0 |
| 12 | Hurricane | 32.7 – 36.9 | 35.0 |

**Table 3**

Conversion relationship between weather type and sunshine duration coefficient.

| Weathertype (**WT1**) | Sunny | Clear to overcast | Cloudy | Overcast | Rainy | Snow | Dust | Haze |
|---|---|---|---|---|---|---|---|---|
| Weathertype (**WT2**) | Sunny | Clear to overcast | Cloudy | Overcast | **Rainy**[a] | **Snow**[b] | Dust | Haze |
| Coeffient (α) | 0.9 | 0.7 | 0.5 | 0.3 | 0.1 | 0.1 | 0.2 | 0.2 |

Note: **Rain**[a] (including light rain, moderate rain, heavy rain, showers, rainstorms, thunderstorms, and sleet);
**Snow**[b] (including light snow, moderate snow, heavy snow, and snow showers).

Given the results of previous studies (Traore et al., 2013; Feng et al., 2017; Feng et al., 2018; Mattar et al., 2018; Fan et al., 2018; Fan et al., 2019; Jiang et al., 2019) and considering the correlation between meteorological variables and $ET_o$ (Landeras et al., 2008; Antonopoulos et al., 2017; Yin et al., 2020; Liu et al., 2022), public weather forecasts include four variables: daily maximum temperature ($T_{max}$), daily minimum temperature ($T_{min}$), wind scale (WS) and weather type (WT). The CatBoost model can directly process feature data. Four different combinations of inputs, C1 (Tmax, Tmin, SDun, and Wspd), C2 ($T_{max}$, $T_{min}$, and SDun), C3 ($T_{max}$, $T_{min}$, and Wspd) and C4

https://doi.org/10.5194/hess-2023-158
Preprint. Discussion started: 31 July 2023
© Author(s) 2023. CC BY 4.0 License.


($T_{max}$ and $T_{min}$), were selected for the four machine learning algorithms, i.e., MLP, XGBoost, LightGBM and CatBoost1. Five different combinations of inputs, C5 ($T_{max}$, $T_{min}$, WT1, and WS), C6 ($T_{max}$, $T_{min}$, WT2, and WS), C7 ($T_{max}$, $T_{min}$, and WT1), C8 ($T_{max}$, $T_{min}$, and WS) and C9 ($T_{max}$, $T_{min}$, and WT2), were selected for the CatBoost2 machine learning algorithm.

The daily observed meteorological data from 1995-2019, public weather forecast data with a 1-day lead time from 2014-2019 and public weather forecast data with a 7-day lead time from 2020-2021 in this study need to be normalized according to Equation (4) before being input into the model (parsing of weather forecast information before data normalization).

$$x^* = \frac{x - \mu}{\sigma} \tag{4}$$

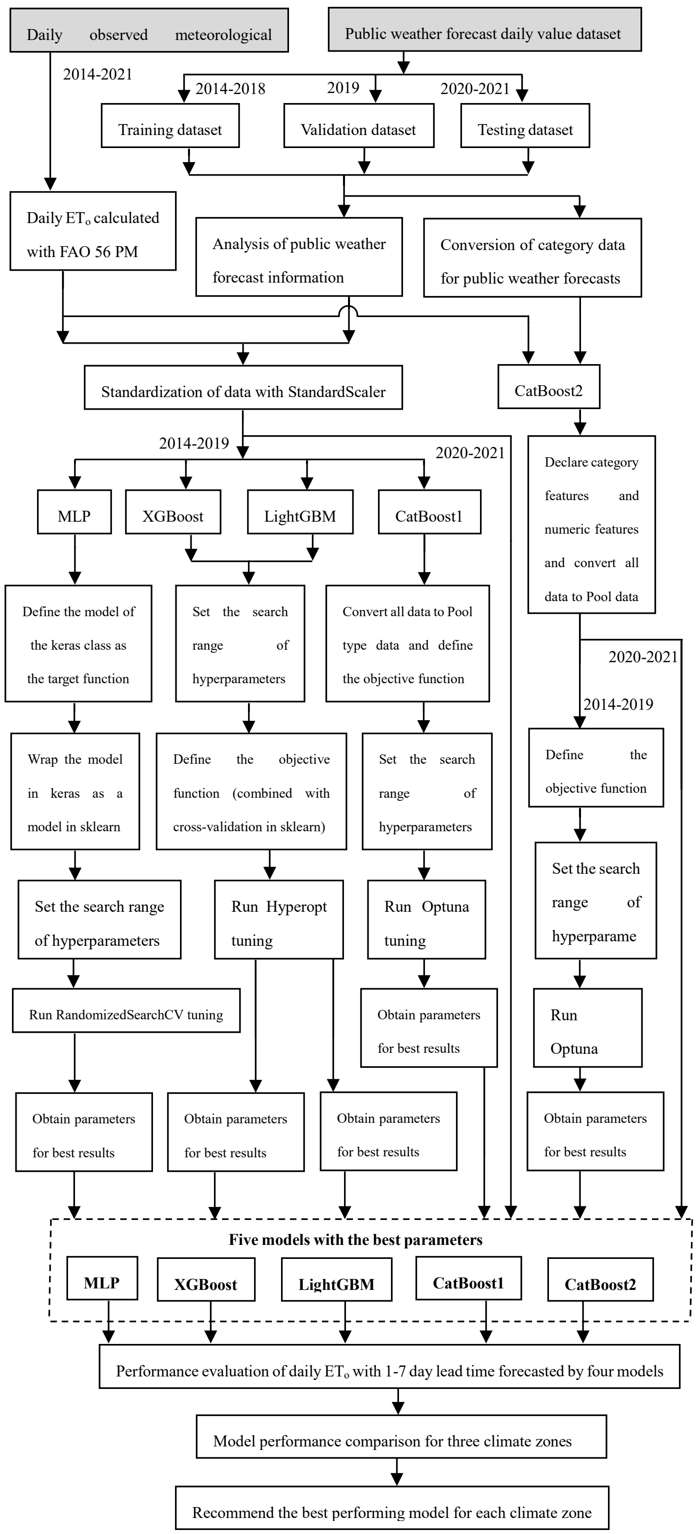
where x* is the x-standardized variable, x is the observed or predicted value of the weather variable, μ is the mean of the sample data, and σ is the variance of the sample data.

It is well known that the performance of machine learning models is directly related to the hyperparameters. The common hyperparameter tuning methods for machine learning models are traditional manual search, grid search (GridSearchCV), randomized search (RandomizedSearchCV) and Bayesian search (BayesSearchCV). In recent years, some new tuning methods have emerged, such as Optuna and Hyperopt, which are two of the more popular hyperparameter tuning tools for machine learning models. Optuna is an automatic hyperparameter optimization framework for automated hyperparameter search that can be used with any machine learning or deep learning framework. Hyperopt is a Python "distributed asynchronous algorithm configuration/ hyperparameter optimization" class library, which is a tool for tuning parameters by Bayesian optimization to perform intelligent searches for optimal parameters of machine learning models. In this study, XGBoost and LightGBM use the Hyperopt method with 5-fold cross-validation in the

https://doi.org/10.5194/hess-2023-158
Preprint. Discussion started: 31 July 2023
© Author(s) 2023. CC BY 4.0 License.


tuning process, CatBoost uses the Optuna method, and MLP uses RandomizedSearchCV with 3-fold cross-validation in the tuning process. Each input combination for each machine learning model was debugged at least three times for comparison to obtain the best hyperparameter combination. The development environment used was Jupyter Notebook 6.0.3, and the following libraries and version information were used: Python 3.7.6, TensorFlow 2.8.0, Scikit-learn 0.22.1, Hyperopt 0.2.7,

https://doi.org/10.5194/hess-2023-158
Preprint. Discussion started: 31 July 2023
© Author(s) 2023. CC BY 4.0 License.
Hydrology and Earth System Sciences
Open Access
Discussions



https://doi.org/10.5194/hess-2023-158
Preprint. Discussion started: 31 July 2023
© Author(s) 2023. CC BY 4.0 License.


**Fig. 2**. Flow of five machine learning methods to predict $ET_o$ in this study

XGBoost 1.5.2, LightGBM 3.3.2, CatBoost 1.0.4, Optuna 2.10.0, NumPy 1.21.5, Pandas 1.0.1, and keras.api._v2.keras 2.8.0. The machine learning models were trained and tested on an Intel(R) Core(TM) i7-10750H CPU with 2.60 GHz-5.0 GHz, 16.0 GB RAM, and an NVIDIA Quadro® P620 graphics card on a graphics workstation. The five machine learning algorithms used to predict $ET_o$ and the parameter tuning process are shown in Figure 2.

## 2.3 Statistical analysis

To evaluate the performance of public weather forecast data and the performance of nine machine learning algorithms for predicting $ET_o$, four statistical indicators, the mean absolute error (MAE), root mean square error (RMSE), the ratio of means(RM), and correlation coefficient (R), were selected. The MAE is the mean of the absolute error and reflects the actual error between the predicted and observed values, the RMSE measures the deviation between the predicted and observed values, and R reflects the degree of correlation between the predicted and observed values. The smaller the values of the MAE and RMSE are, the better, and the closer to 1 the value of R is, the better. The RM is expressed as the ratio of the mean of the predicted value to the mean of the observed value. The RM value can be greater than or less than 1, reflecting the overestimation/underestimation of the predicted value to the observed value, respectively (Tomas-Burguera et al., 2017; Yang et al., 2019a and 2019b). The four statistical indicators are calculated as follows (Mallikarjuna et al. 2014; Kisi and Zounemat-Kermani 2014; Luo et al. 2014; Despotovic et al. 2015; Tomas-Burguera et al. 2017; Yang et al., 2019a and 2019b; Wu et al., 2019b; Zhang et al., 2020):

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n} |P_i - O_i|}{n} \quad (5)$$

https://doi.org/10.5194/hess-2023-158
Preprint. Discussion started: 31 July 2023
© Author(s) 2023. CC BY 4.0 License.


$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(P_i - O_i)^2}{n}} \tag{6}$$

$$\mathrm{RM} = \frac{\overline{P}}{\overline{O}} \tag{7}$$

$$\mathrm{R} = \frac{\sum_{i=1}^{n}(P_i - \overline{P})(O_i - \overline{O})}{\sqrt{\sum_{i=1}^{n}(P_i - \overline{P})^2 \sum_{i=1}^{n}(O_i - \overline{O})^2}} \tag{8}$$

where $P_i$ is the predicted value; $O_i$ is the observed value; i is the sample number, i=1,2,…n; nis the number of samples; $\overline{P}$ is the mean of the predicted value when the number of samples is n; $\overline{O}$ is the mean of the observed value when the number of samples is n;

## 3. Results and discussion

### 3.1. Forecast performance evaluation of weather variables in public weather forecasts

#### 3.1.1. Forecast performance with a single public weather forecast parameter (2014-2021)

The performance statistics of the daily scale forecast weather variables for the three study climate zones (nine stations) obtained from the 1-day ahead public weather forecasts for 2014-2019 are shown in Table 4 and those for 2020-2021 are shown in Figure 3.

For the three climate zones, during the model training period (2014-2018), the mean MAE, RMSE, and R values for $T_{max}$ with a 1-day lead time ranged from 2.53-2.81°C, 3.24-3.58°C, and 0.94-0.96, respectively, and those for $T_{min}$ ranged from 2.09-2.21°C, 2.70-2.83°C, and 0.96-0.97, respectively. The accuracies of the $T_{min}$ forecasts for all three climate zones were higher than those of the $T_{max}$ forecasts. In addition, the mean RM values of $T_{max}$ and $T_{min}$ varied in the ranges of 0.99-1.00 and 0.99-1.11, respectively, indicating that $T_{max}$ was slightly underestimated in all three climate zones, $T_{min}$ was slightly underestimated in both the AR and SAR climate zones and overestimated by 11.39% in the SHZ climate zone (this is mainly due to the poor forecast of $T_{min}$ at the XJ station, resulting in an overestimation of $T_{min}$ by 20.64% at this station). During the model validation period

© Author(s) 2023. CC BY 4.0 License.


(2019), the mean MAE, RMSE, and R values of $T_{max}$ in the three climate zones with a 1-day lead time ranged from 2.59-2.92°C, 3.27-3.83°C, and 0.93-0.95, respectively, and those for $T_{min}$ were 2.18-2.35°C, 2.76-2.97°C, and 0.955-0.962, respectively. The accuracies of the $T_{min}$ forecasts for all three climate zones were higher than those of the $T_{max}$ forecasts. In addition, the mean RM values of $T_{max}$ and $T_{min}$ varied in the ranges of 0.98-0.99 and 1.009-1.014, respectively, indicating that $T_{max}$ was slightly underestimated and $T_{min}$ was slightly overestimated for all three climate zones.

During the model testing period (2020-2021), the mean MAE, RMSE, and R values for $T_{max}$ with a 1-7 day lead time for the three climate zones ranged from 3.78-4.05°C, 4.87-5.18°C, and 0.87-0.91, respectively, and those for $T_{min}$ ranged from 3.05-3.26°C, 3.87-4.09°C and 0.91-0.93, respectively. The accuracies of the $T_{min}$ forecasts for the three climate zones were higher than those of the $T_{max}$ forecasts. The forecast performance of $T_{min}$ and $T_{max}$ decreased with increasing forecasting period. This result is consistent with results from most previous studies in China (Luo et al., 2014 and 2015; Xiong et al., 2016; Yang et al., 2016; Traore et al., 2016; Li et al., 2018; Yang et al. al., 2019a, 2019b; Yin et al., 2020; Liu et al., 2020). In addition, the mean RM values of $T_{max}$ and $T_{min}$ varied in the ranges of 0.98-0.99 and 1.01-1.05, respectively, indicating that $T_{max}$ was slightly underestimated and $T_{min}$ was slightly overestimated at all three sites.

For SDun in the three climate zones, during the model training period (2014-2018), the mean MAE, RMSE, R, and RM values for the 1-day ahead predictions ranged from 2.24-2.38 h, 3.01-3.11 h, 0.68-0.70 and 0.84-0.88, respectively, and SDun was underestimated by 15.61% (AR), 12.51% (SAR), and 11.7% (SHZ). During the model validation period (2019), the mean MAE, RMSE, R, and RM values for the 1-day ahead predictions ranged from 2.09-2.21 h, 2.72-2.92 h, 0.70-0.71 and 0.94-1.03, respectively, and SDun was underestimated by 5.65% and 4.02% for AR and SAR, respectively, and overestimated by 2.82% for SHZ (this is mainly due to the poor forecasts of weather types at the XJ station, resulting in an overestimation of 9.45% for this station). Finally, during the model testing period (2020-2021), the mean MAE, RMSE, R, and RM values for the 1-7 day lead time predictions ranged from 3.36-3.79 h, 4.13-4.62 h, 0.13-0.17, and 0.85-0.96, respectively, and SDun was underestimated by 12.77% (AR), 14.87% (SAR), and 3.86% (SHZ). The SDun forecast performance decreased with increasing forecasting period. This result is consistent with results of previous studies in China (Yang et al., 2016; Yang et al., 2019b; Liu et al., 2020). The poor SDun prediction performance compared with that of the temperature forecast may be due to the large errors in the conversion of weather types from public weather forecasts to SDun (Cai et al., 2007; Yang et al., 2016; Traore et al., 2016; Yang et al., 2019b; Liu et al. et al., 2020).

For the three climate zones, the mean MAE, RMSE, R, and RM values of the 1-day ahead Wspd during the model training period (2014-2018) ranged from 3.57-3.62 m $s^{-1}$, 4.02-4.19 m $s^{-1}$, 0.18-0.24, and 1.51-1.89, respectively, and Wspd was overestimated by 81.56% (AR), 50.69% (SAR) and 89.01% (SHZ). During the model validation period (2019), the mean MAE, RMSE, R and RM values of the 1-day ahead Wspd ranged from 2.77-2.84 m $s^{-1}$, 3.51-3.75 m $s^{-1}$, 0.13-0.15 and 1.65-2.39, respectively, and Wspd was overestimated by 92.81% (AR), 65.38% (SAR) and 138.51% (SHZ). The mean MAE, RMSE, R, and RM values of Wspd with a 1-7 day lead time ranged from 1.28-1.41 m $s^{-1}$, 2.02-2.10 m $s^{-1}$, 0.03-0.06, and 1.18-1.48, respectively, for the three climate zones during the model testing period (2020-2021), and Wspd was overestimated by 47.53% (AR), 17.82% (SAR), and 36.86% (SHZ). It can be seen that the worst predictions among the four variables of the public weather forecasts were those for Wspd, which may be caused by the poor forecasting of wind scale in the public weather forecast and the error in converting wind scale to wind speed (Cai et al.,

https://doi.org/10.5194/hess-2023-158
Preprint. Discussion started: 31 July 2023
© Author(s) 2023. CC BY 4.0 License.
Hydrology and Earth System Sciences
Discussions
Open Access

2007; Yang et al., 2016; Yang et al., 2019a; Liu et al. 2020).

**Table 4**

Statistics for the $T_{max}$, $T_{min}$, Wspd and SDun forecast performance at nine sites in three climate zones for a 1-day lead time during the model training period (2014-2018) and validation period (2019).

| Stage | Station | $T_{max}$ | | | | $T_{min}$ | | | | Wspd | | | | SDun | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | MAE (°C) | RMSE (°C) | R | RM | MAE (°C) | RMSE (°C) | R | RM | MAE ($ms^{-1}$) | RMSE ($ms^{-1}$) | R | RM | MAE (h) | RMSE (h) | R | RM |
| **Training** | AR/HN | 2.36 | 3.04 | 0.97 | 0.99 | 2.33 | 3.03 | 0.97 | 0.95 | 3.63 | 4.17 | 0.17 | 1.68 | 2.33 | 3.15 | 0.68 | 0.81 |
| | AR/YC | 2.39 | 3.03 | 0.97 | 1.00 | 2.03 | 2.60 | 0.97 | 1.00 | 3.68 | 4.39 | 0.13 | 2.22 | 2.13 | 2.79 | 0.70 | 0.92 |
| | AR/ZW | 2.65 | 3.39 | 0.95 | 0.99 | 2.28 | 2.91 | 0.96 | 1.03 | 3.56 | 4.03 | 0.24 | 1.55 | 2.27 | 3.04 | 0.70 | 0.82 |
| | AR/ZN | 2.73 | 3.49 | 0.95 | 1.00 | 2.17 | 2.77 | 0.97 | 0.97 | 3.63 | 4.17 | 0.17 | 1.68 | 2.33 | 3.15 | 0.68 | 0.81 |
| | **Average** | **2.53** | **3.24** | **0.96** | **1.00** | **2.20** | **2.83** | **0.97** | **0.99** | **3.62** | **4.19** | **0.19** | **1.82** | **2.27** | **3.01** | **0.69** | **0.84** |
| | SAR/YAC | 2.71 | 3.46 | 0.95 | 0.99 | 2.55 | 3.26 | 0.96 | 1.00 | 3.63 | 4.18 | 0.22 | 1.81 | 2.15 | 2.89 | 0.71 | 0.88 |
| | SAR/HY | 2.81 | 3.60 | 0.94 | 0.99 | 1.76 | 2.25 | 0.97 | 1.02 | 3.56 | 4.02 | 0.25 | 1.55 | 2.22 | 2.98 | 0.70 | 0.92 |
| | SAR/TX | 2.90 | 3.68 | 0.94 | 0.99 | 1.98 | 2.58 | 0.97 | 0.93 | 3.51 | 3.87 | 0.25 | 1.17 | 2.35 | 3.18 | 0.69 | 0.83 |
| | **Average** | **2.81** | **3.58** | **0.94** | **0.99** | **2.09** | **2.70** | **0.97** | **0.99** | **3.57** | **4.02** | **0.24** | **1.51** | **2.24** | **3.01** | **0.70** | **0.88** |
| | SHZ/GY | 2.68 | 3.44 | 0.94 | 0.99 | 1.93 | 2.42 | 0.97 | 1.02 | 3.57 | 4.05 | 0.20 | 1.62 | 2.40 | 3.13 | 0.83 | 0.69 |
| | SHZ/XJ | 2.44 | 3.13 | 0.95 | 0.99 | 2.49 | 3.10 | 0.95 | 1.21 | 3.58 | 4.29 | 0.17 | 2.16 | 2.36 | 3.09 | 0.67 | 0.93 |
| | **Average** | **2.56** | **3.28** | **0.94** | **0.99** | **2.21** | **2.76** | **0.96** | **1.11** | **3.58** | **4.17** | **0.18** | **1.89** | **2.38** | **3.11** | **0.68** | **0.88** |
| **Validation** | AR/HN | 2.59 | 3.34 | 0.96 | 0.99 | 2.58 | 3.21 | 0.96 | 1.01 | 2.75 | 3.57 | 0.13 | 1.82 | 2.02 | 2.63 | 0.73 | 0.90 |
| | AR/YC | 2.56 | 3.34 | 0.96 | 0.99 | 2.11 | 2.75 | 0.97 | 1.00 | 2.81 | 3.71 | 0.08 | 2.23 | 2.01 | 2.67 | 0.71 | 0.99 |
| | AR/ZW | 2.80 | 3.69 | 0.94 | 1.00 | 2.37 | 2.99 | 0.96 | 1.03 | 2.76 | 3.52 | 0.19 | 1.73 | 2.02 | 2.67 | 0.72 | 0.94 |
| | AR/ZN | 2.83 | 3.71 | 0.94 | 0.99 | 2.33 | 2.94 | 0.96 | 1.00 | 2.78 | 3.59 | 0.16 | 1.93 | 2.30 | 2.90 | 0.65 | 0.95 |
| | **Average** | **2.69** | **3.52** | **0.95** | **0.99** | **2.35** | **2.97** | **0.96** | **1.01** | **2.77** | **3.60** | **0.14** | **1.93** | **2.09** | **2.72** | **0.71** | **0.94** |
| | SAR/YAC | 2.81 | 3.67 | 0.94 | 0.99 | 2.73 | 3.38 | 1.05 | 0.95 | 2.90 | 3.73 | 0.13 | 2.03 | 1.99 | 2.58 | 0.75 | 1.01 |
| | SAR/HY | 2.94 | 3.88 | 0.92 | 0.98 | 1.83 | 2.37 | 0.97 | 0.99 | 2.71 | 3.48 | 0.12 | 1.63 | 2.19 | 2.98 | 0.67 | 0.95 |
| | SAR/TX | 3.00 | 3.94 | 0.93 | 0.98 | 1.99 | 2.55 | 0.97 | 0.98 | 2.71 | 3.33 | 0.14 | 1.31 | 2.10 | 2.74 | 0.92 | 0.92 |
| | **Average** | **2.92** | **3.83** | **0.93** | **0.98** | **2.18** | **2.77** | **0.96** | **1.01** | **2.78** | **3.51** | **0.13** | **1.65** | **2.10** | **2.77** | **0.72** | **0.96** |
| | SHZ/GY | 2.66 | 3.37 | 0.94 | 0.99 | 2.04 | 2.54 | 0.96 | 1.05 | 2.78 | 3.65 | 0.16 | 2.05 | 2.27 | 2.99 | 0.69 | 0.96 |
| | SHZ/XJ | 2.52 | 3.17 | 0.94 | 0.97 | 2.38 | 2.98 | 0.95 | 0.98 | 2.89 | 3.84 | 0.14 | 2.72 | 2.16 | 2.85 | 0.72 | 1.09 |
| | **Average** | **2.59** | **3.27** | **0.94** | **0.98** | **2.21** | **2.76** | **0.96** | **1.01** | **2.84** | **3.75** | **0.15** | **2.39** | **2.21** | **2.92** | **0.70** | **1.03** |

https://doi.org/10.5194/hess-2023-158
Preprint. Discussion started: 31 July 2023
© Author(s) 2023. CC BY 4.0 License.

Hydrology and Earth System Sciences Discussions

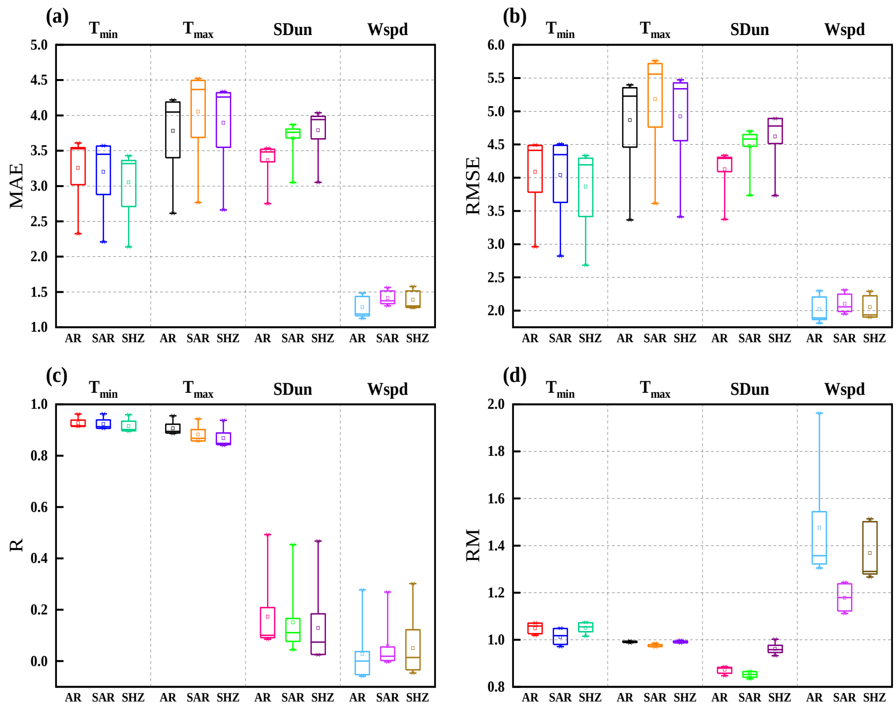



**Fig. 3.** Statistics of the forecast performance indices for daily scale forecasts of $T_{max}$, $T_{min}$, SDun and Wspd obtained from public weather forecasts with 1-7 day lead times in three climate regions (AR, SAR and SHZ) in 2020-2021. ((a) MAE, (b) RMSE, (c) RM, and (d) R).

### 3.1.2. Seasonality of the weather variable forecast performance metrics with public weather forecasts (2020-2021)

The mean predicted performance metrics of weather variables for each season using public weather forecasts 1-7 days ahead for the three study sites in 2020-2021 are shown in Tables 5-7. The average performance metrics for $T_{max}$ in the AR (SAR) climate zone all decreased sequentially in the following order: summer, fall, winter, and spring. The mean MAE, RMSE, and R values for $T_{max}$ ranged from 3.373-4.610°C, 4.244-5.733°C, and 0.399-0.777 (3.532-4.841°C, 4.505-5.977°C, and 0.387-0.743), respectively. The average performance metrics for $T_{max}$ in the SHZ climate zone all decreased in the following order: fall, summer, winter, and spring. The mean MAE, RMSE, and R values for $T_{max}$ varied in the ranges of 3.526-4.592°C, 4.464-5.657°C, and 0.253-0.755,

https://doi.org/10.5194/hess-2023-158
Preprint. Discussion started: 31 July 2023
© Author(s) 2023. CC BY 4.0 License.


respectively. The $T_{max}$ values for the three climate zones were overestimated by 2.93%-4.65% (1.10%-1.38%) for spring (summer) (except for the SAR climate zone, where $T_{max}$ was underestimated by 0.24% for summer) and underestimated by 4.30%-6.33% (31.18%-53.20%) for fall (winter).

The mean performance metrics for $T_{min}$ in the AR (SAR) climate zone all decreased sequentially in the following order: fall, summer, winter, and spring. The mean MAE, RMSE, and R values for $T_{min}$ ranged from 2.952-3.717°C, 3.672-4.630°C, 0.610-0.846 (2.792-3.674°C, 3.513-4.520 °C, 0.617-0.849), respectively. The average performance metrics for $T_{min}$ in the SHZ climate zone all decreased in the following order: summer, fall, winter, and spring, and the mean MAE, RMSE, and R values of $T_{min}$ varied in the ranges of 2.588-3.445 °C, 3.357-4.320 °C, and 0.596-0.811, respectively. The $T_{min}$ values were overestimated by 4.08%-54.4% (4.43%-5.68% and 0.40%-3.25%) in spring (summer and winter, respectively) for the three climate zones (except for the $T_{min}$ value in winter in the AR climate zone, which was underestimated by 0.05%) and underestimated by 6.65%-10.10% in fall.

The mean performance metrics for SDun in all three climate zones decreased in the following order: winter, spring, fall, and summer, and the mean MAE, RMSE and R values in the AR and SAR (SHZ) climate zones varied in the ranges of 2.348-3.866 h, 2.822-4.715 h, 0.058-0.272 and 2.585-4.260 h, 3.077-5.126 h and 0.040-0.213 (2.683-4.279 h, 3.160-5.208 h, and 0.033-0.323), respectively. SDun was underestimated by 11.45%-16.43%, 1.70%-14.97%, 10.07%-10.8% and 7.65%-9.93% in spring, summer, fall and winter for the three climatic zones, respectively, except in the SHZ climatic zone, where SDun was overestimated by 12.80% in fall.

The mean performance metrics for Wspd in the AR (SAR) climate zone all decreased in the

© Author(s) 2023. CC BY 4.0 License.


following order: fall, summer, spring, and winter, and the mean MAE, RMSE, and R values of Wspd varied in the ranges of 1.180-1.467 m $s^{-1}$, 1.818-2.345 m $s^{-1}$, and 0.017-0.044 (1.267-1.629 m $s^{-1}$, 1.885-2.501 m $s^{-1}$, 0.030-0.063), respectively. The mean performance metrics for Wspd in the SHZ climate zone decreased in the following order: summer, fall, spring, and winter, and the mean MAE, RMSE, and R values varied in the ranges of 1.255-1.571 m $s^{-1}$, 1.829-2.371 m $s^{-1}$, 0.040-0.062 0.040-0.062. Wspd was overestimated by 17.30%-40.63%, 2.43%-25.78%, 24.37%-50.23%, and 40.50%-60.30% in spring, summer, fall, and winter for the three climatic zones, respectively.

In conclusion, when considering all the metrics, the prediction performance for $T_{max}$ and $T_{min}$ were best in fall and summer, followed by winter and spring. In contrast, the prediction performance for SDun was consistent across all three climate zones, with the best prediction performance occurring in winter and spring, followed by fall and summer. The prediction performance for Wspd showed great variation depending on the climate zone.

**Table 5**

Seasonal average statistics of performance indicators for 1-7 day lead time weather variables ($T_{max}$, $T_{min}$, Wspd and SDun) predicted by public weather forecasts at 4 stations in the AR climate zone in 2020-2021

| Stations/Seasons | $T_{max}$ | | | | $T_{min}$ | | | | Wspd | | | | SDun | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | MAE (°C) | RMSE (°C) | RM | R | MAE (°C) | RMSE (°C) | RM | R | MAE (m $s^{-1}$) | RMSE (m $s^{-1}$) | RM | R | MAE (h) | RMSE (h) | RM | R |
| **AR/Huinong** | | | | | | | | | | | | | | | | |
| Spring | 4.399 | 5.462 | 1.031 | 0.581 | 3.808 | 4.722 | 0.462 | 0.710 | 1.541 | 2.319 | 1.317 | 0.008 | 3.447 | 4.053 | 0.797 | 0.174 |
| Summer | 3.291 | 4.114 | 1.008 | 0.438 | 2.763 | 3.480 | 1.038 | 0.689 | 1.214 | 1.863 | 1.108 | 0.006 | 3.908 | 4.769 | 0.817 | 0.040 |
| Fall | 3.254 | 4.257 | 0.953 | 0.817 | 3.076 | 3.829 | 0.936 | 0.853 | 1.112 | 1.749 | 1.277 | 0.025 | 3.860 | 4.620 | 0.812 | 0.052 |
| Winter | 3.791 | 4.898 | 0.603 | 0.696 | 3.817 | 4.628 | 1.004 | 0.696 | 1.618 | 2.413 | 1.444 | 0.040 | 2.446 | 2.773 | 0.792 | 0.351 |
| **AR/Yinchuan** | | | | | | | | | | | | | | | | |
| Spring | 4.412 | 5.501 | 1.038 | 0.569 | 3.627 | 4.520 | 1.857 | 0.708 | 1.263 | 2.115 | 1.620 | 0.022 | 3.350 | 3.982 | 0.855 | 0.173 |
| Summer | 3.304 | 4.135 | 1.013 | 0.441 | 2.827 | 3.602 | 1.056 | 0.664 | 1.078 | 1.754 | 1.596 | 0.003 | 3.674 | 4.505 | 0.848 | 0.074 |
| Fall | 3.300 | 4.339 | 0.955 | 0.807 | 2.680 | 3.361 | 0.959 | 0.872 | 1.284 | 1.899 | 1.988 | -0.02 | 3.662 | 4.419 | 0.910 | 0.056 |
| Winter | 3.681 | 4.817 | 0.647 | 0.700 | 3.527 | 4.346 | 0.995 | 0.704 | 1.457 | 2.340 | 1.973 | -0.02 | 2.391 | 2.904 | 0.953 | 0.332 |
| **AR/Zhongwei** | | | | | | | | | | | | | | | | |
| Spring | 4.804 | 5.971 | 1.038 | 0.497 | 3.831 | 4.793 | 0.608 | 0.709 | 1.511 | 2.255 | 1.211 | 0.062 | 3.333 | 3.976 | 0.883 | 0.162 |
| Summer | 3.394 | 4.282 | 1.021 | 0.350 | 3.278 | 4.154 | 1.073 | 0.546 | 1.285 | 1.809 | 1.014 | 0.037 | 3.987 | 4.813 | 0.905 | 0.080 |
| Fall | 3.619 | 4.727 | 0.955 | 0.741 | 3.071 | 3.780 | 0.918 | 0.825 | 1.110 | 1.744 | 1.182 | 0.073 | 3.730 | 4.528 | 0.901 | 0.054 |

© Author(s) 2023. CC BY 4.0 License.


| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Winter | 4.200 | 5.365 | 0.699 | 0.645 | 3.414 | 4.277 | 0.988 | 0.713 | 1.410 | 2.299 | 1.429 | 0.047 | 2.362 | 2.881 | 0.936 | 0.218 |
| **AR/Zhongning** | | | | | | | | | | | | | | | | |
| Spring | 4.823 | 5.998 | 1.041 | 0.494 | 3.603 | 4.484 | 1.236 | 0.699 | 1.446 | 2.301 | 1.477 | 0.026 | 3.108 | 3.782 | 0.895 | 0.155 |
| Summer | 3.503 | 4.444 | 1.013 | 0.366 | 3.309 | 4.182 | 1.060 | 0.541 | 1.183 | 1.859 | 1.313 | 0.022 | 3.894 | 4.773 | 0.892 | 0.064 |
| Fall | 3.702 | 4.873 | 0.952 | 0.743 | 2.982 | 3.716 | 0.921 | 0.832 | 1.212 | 1.880 | 1.562 | 0.097 | 3.630 | 4.440 | 0.945 | 0.068 |
| Winter | 4.033 | 5.275 | 0.804 | 0.647 | 3.244 | 4.007 | 1.011 | 0.715 | 1.383 | 2.327 | 1.566 | 0.015 | 2.194 | 2.728 | 0.932 | 0.187 |
| **AR/Average** | | | | | | | | | | | | | | | | |
| Spring | 4.610 | 5.733 | 1.037 | 0.535 | 3.717 | 4.630 | 1.041 | 0.707 | 1.440 | 2.248 | 1.406 | 0.030 | **3.310** | **3.948** | **0.858** | **0.166** |
| Summer | **3.373** | **4.244** | **1.014** | **0.399** | **3.044** | **3.855** | **1.057** | **0.610** | **1.190** | **1.821** | **1.258** | **0.017** | 3.866 | 4.715 | 0.866 | 0.065 |
| Fall | **3.469** | **4.549** | **0.954** | **0.777** | **2.952** | **3.672** | **0.934** | **0.846** | **1.180** | **1.818** | **1.502** | **0.044** | 3.721 | 4.502 | 0.892 | 0.058 |
| Winter | 3.926 | 5.089 | 0.688 | 0.672 | 3.501 | 4.315 | 1.000 | 0.707 | 1.467 | 2.345 | 1.603 | 0.021 | **2.348** | **2.822** | **0.903** | **0.272** |

Note: The best statistical indicators of each weather variable predicted by public weather forecasts in each climate zone in the four seasons are highlighted in blue, and the better statistical indicators are highlighted in grey.

**Table 6**

Seasonal average statistics of performance indicators for 1-7 day lead time weather variables ($T_{max}$, $T_{min}$, Wspd and SDun) predicted by public weather forecasts at 4 stations in the SAR climate zone in 2020-2021

| Stations/Seasons | $T_{max}$ | | | | $T_{min}$ | | | | Wspd | | | | SDun | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | MAE (°C) | RMSE (°C) | RM | R | MAE (°C) | RMSE (°C) | RM | R | MAE (m s$^{-1}$) | RMSE (m s$^{-1}$) | RM | R | MAE (h) | RMSE (h) | RM | R |
| **SAR/Yanchi** | | | | | | | | | | | | | | | | |
| Spring | 4.778 | 5.918 | 1.014 | 0.496 | 3.988 | 4.997 | 1.092 | 0.664 | 1.538 | 2.390 | 1.398 | 0.002 | 3.716 | 4.357 | 0.777 | 0.122 |
| Summer | 3.454 | 4.379 | 0.990 | 0.415 | 3.313 | 4.183 | 1.035 | 0.607 | 1.070 | 1.680 | 1.131 | 0.024 | 4.384 | 5.181 | 0.726 | 0.044 |
| Fall | 3.544 | 4.662 | 0.930 | 0.779 | 3.305 | 4.124 | 0.879 | 0.825 | 1.121 | 1.797 | 1.496 | 0.060 | 3.909 | 4.694 | 0.789 | 0.036 |
| Winter | 4.188 | 5.447 | 0.052 | 0.680 | 4.476 | 5.420 | 1.023 | 0.650 | 1.759 | 2.722 | 1.706 | 0.095 | 2.542 | 3.014 | 0.890 | 0.275 |
| **SAR/Haiyuan** | | | | | | | | | | | | | | | | |
| Spring | 4.853 | 5.971 | 1.040 | 0.437 | 3.397 | 4.211 | 2.216 | 0.645 | 1.541 | 2.227 | 1.070 | 0.046 | 3.790 | 4.528 | 0.849 | 0.109 |
| Summer | 3.573 | 4.553 | 1.002 | 0.368 | 2.579 | 3.371 | 1.055 | 0.594 | 1.312 | 1.865 | 1.006 | 0.050 | 4.167 | 5.063 | 0.937 | 0.030 |
| Fall | 3.752 | 4.831 | 0.936 | 0.717 | 2.409 | 3.068 | 0.909 | 0.864 | 1.080 | 1.762 | 1.194 | 0.052 | 4.166 | 4.955 | 0.999 | 0.034 |
| Winter | 4.494 | 5.652 | 0.614 | 0.567 | 3.371 | 4.149 | 1.002 | 0.672 | 1.538 | 2.438 | 1.361 | 0.065 | 2.703 | 3.194 | 0.937 | 0.168 |
| **SAR/Tongxin** | | | | | | | | | | | | | | | | |
| Spring | 4.891 | 6.043 | 1.034 | 0.477 | 3.524 | 4.352 | 1.324 | 0.751 | 1.748 | 2.377 | 1.051 | 0.042 | 3.476 | 4.178 | 0.881 | 0.151 |
| Summer | 3.569 | 4.582 | 1.001 | 0.377 | 2.572 | 3.303 | 1.043 | 0.650 | 1.686 | 2.120 | 0.936 | 0.066 | 4.230 | 5.134 | 0.888 | 0.078 |
| Fall | 3.783 | 4.913 | 0.944 | 0.733 | 2.663 | 3.347 | 0.909 | 0.857 | 1.600 | 2.095 | 1.041 | 0.049 | 4.169 | 4.954 | 0.910 | 0.051 |
| Winter | 4.498 | 5.713 | 0.738 | 0.609 | 3.174 | 3.917 | 0.987 | 0.757 | 1.590 | 2.344 | 1.148 | 0.028 | 2.511 | 3.022 | 0.875 | 0.195 |
| **SAR/Average** | | | | | | | | | | | | | | | | |
| Sping | 4.841 | 5.977 | 1.029 | 0.470 | 3.636 | 4.520 | 1.544 | 0.687 | 1.609 | 2.331 | 1.173 | 0.030 | **3.661** | **4.354** | **0.836** | **0.127** |
| Summer | **3.532** | **4.505** | **0.998** | **0.387** | **2.821** | **3.619** | **1.044** | **0.617** | **1.356** | **1.888** | **1.024** | **0.047** | 4.260 | 5.126 | 0.850 | 0.051 |
| Fall | **3.693** | **4.802** | **0.937** | **0.743** | **2.792** | **3.513** | **0.899** | **0.849** | **1.267** | **1.885** | **1.244** | **0.054** | 4.081 | 4.868 | 0.899 | 0.040 |
| Winter | 4.393 | 5.604 | 0.468 | 0.619 | 3.674 | 4.495 | 1.004 | 0.693 | 1.629 | 2.501 | 1.405 | 0.063 | **2.585** | **3.077** | **0.901** | **0.213** |

Note: The best statistical indicators of each weather variable predicted by public weather forecasts in each climate zone in the four seasons are highlighted in blue, and the better statistical indicators are highlighted in grey.

**Table 7**

Seasonal average statistics of performance indicators for 1-7 day lead time weather variables ($T_{max}$, $T_{min}$, Wspd and SDun) predicted by public weather forecasts at 4 stations in the SHZ climate zone in 2020-2021

https://doi.org/10.5194/hess-2023-158
Preprint. Discussion started: 31 July 2023
© Author(s) 2023. CC BY 4.0 License.
Hydrology and Earth System Sciences
Open Access
EGU
Discussions

| Stations/Seasons | $T_{max}$ | | | | $T_{min}$ | | | | Wspd | | | | SDun | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | MAE (°C) | RMSE (°C) | RM | R | MAE (°C) | RMSE (°C) | RM | R | MAE (m s$^{-1}$) | RMSE (m s$^{-1}$) | RM | R | MAE (h) | RMSE (h) | RM | R |
| **SHZ/Guyuan** | | | | | | | | | | | | | | | | |
| Spring | 4.854 | 5.922 | 1.063 | 0.432 | 3.149 | 3.958 | 1.331 | 0.652 | 1.700 | 2.262 | 0.889 | 0.044 | 3.895 | 4.635 | 0.819 | 0.101 |
| Summer | 3.783 | 4.733 | 1.029 | 0.308 | 2.318 | 3.054 | 1.058 | 0.607 | 1.428 | 1.886 | 0.890 | 0.058 | 4.287 | 5.207 | 0.965 | 0.055 |
| Fall | 3.661 | 4.618 | 0.970 | 0.730 | 2.397 | 3.037 | 0.916 | 0.854 | 1.329 | 1.864 | 0.960 | 0.049 | 4.265 | 5.055 | 1.038 | 0.065 |
| Winter | 4.381 | 5.506 | 0.428 | 0.557 | 3.295 | 4.058 | 1.040 | 0.674 | 1.638 | 2.285 | 0.962 | 0.083 | 2.582 | 3.034 | 0.833 | 0.378 |
| **SHZ/Xiji** | | | | | | | | | | | | | | | | |
| Spring | 4.330 | 5.392 | 1.030 | 0.467 | 3.741 | 4.681 | 0.781 | 0.663 | 1.421 | 2.288 | 1.852 | 0.042 | 3.624 | 4.398 | 0.952 | 0.145 |
| Summer | 3.457 | 4.411 | 0.993 | 0.198 | 2.857 | 3.659 | 1.055 | 0.585 | 1.081 | 1.771 | 1.578 | 0.032 | 4.270 | 5.208 | 1.001 | 0.010 |
| Fall | 3.390 | 4.309 | 0.944 | 0.779 | 3.186 | 3.889 | 0.893 | 0.767 | 1.262 | 1.943 | 1.870 | 0.075 | 4.110 | 4.919 | 1.218 | 0.077 |
| Winter | 3.668 | 4.641 | 0.537 | 0.577 | 3.507 | 4.363 | 1.025 | 0.714 | 1.503 | 2.457 | 2.036 | -0.004 | 2.784 | 3.285 | 1.014 | 0.268 |
| **SHZ/Average** | | | | | | | | | | | | | | | | |
| Sping | 4.592 | 5.657 | 1.047 | 0.450 | 3.445 | 4.320 | 1.056 | 0.658 | 1.561 | 2.275 | 1.371 | 0.043 | **3.760** | **4.517** | **0.886** | **0.123** |
| Summer | **3.620** | **4.572** | **1.011** | **0.253** | **2.588** | **3.357** | **1.057** | **0.596** | **1.255** | **1.829** | **1.234** | **0.045** | 4.279 | 5.208 | 0.983 | 0.033 |
| Fall | **3.526** | **4.464** | **0.957** | **0.755** | **2.792** | **3.463** | **0.905** | **0.811** | **1.296** | **1.904** | **1.415** | **0.062** | 4.188 | 4.987 | 1.128 | 0.071 |
| Winter | 4.025 | 5.074 | 0.483 | 0.567 | 3.401 | 4.211 | 1.033 | 0.694 | 1.571 | 2.371 | 1.499 | 0.040 | **2.683** | **3.160** | **0.924** | **0.323** |

Note: The best statistical indicators of each weather variable predicted by public weather forecasts in each climate zone in the four seasons are highlighted in blue, and the better statistical indicators are highlighted in grey.

### 3.2. Performance comparison of four machine learning models trained and validated based on daily observed meteorological data to predict $ET_o$ with various input combinations

The predicted daily $ET_o$ performance statistics of the four machine learning models based on daily observed meteorological data trained (1995-2015) and validated (2016-2019) with different input combinations for the three climate zones, AR, SAR and SHZ, are shown in Tables 8-10, respectively. The predicted daily $ET_o$ performance was dependent on the machine learning model type, input combination and climate zone and significantly differed.

For the AR climate zone, $MLP_o$ and $XGBoost_o$ had the best prediction performance with input combination C1, and $XGBoost_o$ and $LightGBM_o$ had the best prediction performance with input combinations C2, C3 and C4 during the training and validation periods. During the testing period, the performance of $MLP_o$, $XGBoost_o$ and $CatBoost1_o$ in predicting $ET_o$ decreased, with the use of input combination C2 yielding the largest decrease, followed by combinations C4, C1, and C3, and

https://doi.org/10.5194/hess-2023-158
Preprint. Discussion started: 31 July 2023
© Author(s) 2023. CC BY 4.0 License.


for $LightGBM_o$ the order of decreasing performance was as follows: C4, C2, C1, and C3. $LightGBM_o$ was the best model for predicting $ET_o$ with input combination C3 (MAE = 0.837 mm $d^{-1}$, RMSE = 1.113 mm $d^{-1}$, and R = 0.826), and $CatBoost1_o$ was the best model in terms of prediction performance with input combinations C1, C2, and C4 (MAE range: 0.770-0.824 mm $d^{-1}$, RMSE range: 1.042-1.084 mm $d^{-1}$, and R range: 0.826-0.849). When using input combination C2, $CatBoost1_o$ was the best performing machine learning model in the AR climate zone testing period (MAE=0.770 mm $d^{-1}$, RMSE=1.042 mm $d^{-1}$, and R=0.843).

For the SAR climate zone, $XGBoost_o$ and $LightGBM_o$ had the best prediction performance with various input combinations during the training and validation periods. During the testing period, the performance of $MLP_o$ and $CatBoost1_o$ in predicting $ET_o$ with different input combinations decreased. The use of input combination C4 yielded the largest decrease, followed by combinations C3, C2, and C1, and for $XGBoost_o$ and $LightGBM_o$, the order of decreasing performance was C4, C2, C3, and C1. $LightGBM_o$ was the best performing model in terms of prediction performance (MAE=0.935 mm $d^{-1}$, RMSE=1.261 mm $d^{-1}$, and R=0.787) with the C1 input combination, $XGBoost_o$ was the best performing model in terms of prediction performance (MAE=0.923 mm $d^{-1}$, RMSE=1.267 mm $d^{-1}$, and R=0.791) with the C2 input combination, and $CatBoost1_o$ was the best performing model with both the C3 and C4 input combinations (MAE range: 0.886-0.904 mm $d^{-1}$, RMSE range: 1.204-1.208 mm $d^{-1}$, and R range: 0.796-0.798). $CatBoost1_o$ with the C4 input combination was the best performing model for the SAR climate zone and the best performing machine learning model in the testing period (MAE=0.886 mm $d^{-1}$, RMSE=1.208 mm $d^{-1}$, and R=0.798).

For the SHZ climate zone, $XGBoost_o$ and $LightGBM_o$ had the best prediction performance with

https://doi.org/10.5194/hess-2023-158
Preprint. Discussion started: 31 July 2023
© Author(s) 2023. CC BY 4.0 License.


various input combinations during the training and validation periods. During the testing period, the performance of $MLP_o$ and $XGBoost_o$ in predicting $ET_o$ with different input combinations decreased. The use of input combination C2 yielded the largest performance decrease, followed by combinations C4, C1, and C3. For $LightGBM_o$ the order of decreasing performance was C4, C2, C1, and C3 and that for $CatBoost1_o$ was C1, C2, C4, and C3. The prediction performance of $CatBoost1_o$ with the C1, C2, C3, and C4 input combinations (MAE range: 0.772-0.820 mm $d^{-1}$, RMSE range: 1.044-1.088 mm $d^{-1}$, and R range: 0.741-0.750) was the best. $CatBoost1_o$ with the C1 input combination was the best performing machine learning model in the SAR climate zone testing period (MAE = 0.786 mm $d^{-1}$, RMSE = 1.044 mm $d^{-1}$, and R = 0.750).

**Table 8**

Average statistics of the predictions with a lead time of 1-7 days by the $MLP_o$, $XGBoost_o$, $LightGBM_o$ and $CatBoost1_o$ models for the AR climate zone with different input combinations during training, validation and testing.

| Inputs/model | training | | | | validation | | | | testing | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | MAE (mm/d) | RMSE (mm/d) | RM | R | MAE (mm/d) | RMSE (mm/d) | RM | R | MAE (mm/d) | RMSE (mm/d) | RM | R |
| **$T_{max}$, $T_{min}$, SDun, Wspd** | | | | | | | | | | | | |
| $MLP_o$ | 0.349 | 0.460 | 0.993 | 0.974 | 0.345 | 0.457 | 0.968 | 0.971 | 0.884 | 1.208 | 0.991 | 0.793 |
| $XGBoost_o$ | 0.309 | 0.407 | 1.003 | 0.979 | 0.288 | 0.385 | 0.986 | 0.979 | 0.817 | 1.100 | 0.995 | 0.815 |
| $LightGBM_o$ | 0.386 | 0.500 | 1.002 | 0.970 | 0.357 | 0.467 | 0.991 | 0.968 | 0.840 | 1.115 | 0.986 | 0.820 |
| $CatBoost1_o$ | 0.421 | 0.549 | 1.001 | 0.963 | 0.393 | 0.517 | 0.992 | 0.961 | **0.824** | **1.084** | **0.969** | **0.826** |
| **$T_{max}$, $T_{min}$, SDun** | | | | | | | | | | | | |
| $MLP_o$ | 0.509 | 0.685 | 0.988 | 0.940 | 0.478 | 0.639 | 1.014 | 0.943 | 0.792 | 1.081 | 0.944 | 0.837 |
| $XGBoost_o$ | 0.493 | 0.671 | 0.975 | 0.945 | 0.440 | 0.588 | 1.005 | 0.949 | 0.777 | 1.052 | 0.949 | 0.838 |
| $LightGBM_o$ | 0.510 | 0.677 | 0.995 | 0.941 | 0.472 | 0.615 | 1.024 | 0.945 | 0.793 | 1.058 | 0.959 | 0.837 |
| $CatBoost1_o$ | 0.522 | 0.697 | 0.995 | 0.937 | 0.486 | 0.636 | 1.028 | 0.942 | **0.770** | **1.042** | **0.956** | **0.843** |
| **$T_{max}$, $T_{min}$, Wspd** | | | | | | | | | | | | |
| $MLP_o$ | 0.501 | 0.673 | 0.982 | 0.946 | 0.491 | 0.665 | 0.988 | 0.944 | 0.931 | 1.260 | 1.106 | 0.799 |
| $XGBoost_o$ | 0.485 | 0.651 | 0.986 | 0.947 | 0.445 | 0.593 | 1.010 | 0.950 | **0.837** | **1.113** | **1.068** | **0.826** |
| $LightGBM_o$ | 0.497 | 0.652 | 0.999 | 0.946 | 0.468 | 0.618 | 1.007 | 0.944 | 0.872 | 1.168 | 1.100 | 0.823 |
| $CatBoost1_o$ | 0.558 | 0.721 | 0.998 | 0.934 | 0.522 | 0.679 | 1.010 | 0.932 | 0.862 | 1.147 | 1.101 | 0.830 |
| **$T_{max}$, $T_{min}$** | | | | | | | | | | | | |
| $MLP_o$ | 0.658 | 0.862 | 1.012 | 0.904 | 0.652 | 0.850 | 1.075 | 0.904 | 0.837 | 1.127 | 1.092 | 0.838 |
| $XGBoost_o$ | 0.664 | 0.867 | 0.977 | 0.905 | 0.614 | 0.799 | 1.035 | 0.908 | 0.791 | 1.057 | 1.056 | 0.844 |
| $LightGBM_o$ | 0.699 | 0.892 | 0.991 | 0.902 | 0.646 | 0.818 | 1.050 | 0.904 | 0.805 | 1.046 | 1.063 | 0.844 |
| $CatBoost1_o$ | 0.681 | 0.883 | 0.990 | 0.899 | 0.640 | 0.826 | 1.052 | 0.902 | **0.782** | **1.051** | **1.063** | **0.849** |

© Author(s) 2023. CC BY 4.0 License.


Note: The statistical indicators of the best performing machine learning models with different input combinations for this climate zone are highlighted in blue, and the statistical indicators of the best performing machine learning models with the same input combination for this climate zone are highlighted in grey.

**Table 9**

Average statistics of the predictions with a lead time of 1-7 days by the $MLP_o$, $XGBoost_o$, $LightGBM_o$ and $CatBoost1_o$ models for the SAR climate zone with different input combinations during training, validation and testing.

| Inputs/model | training | | | | validation | | | | testing | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | MAE (mm/d) | RMSE (mm/d) | RM | R | MAE (mm/d) | RMSE (mm/d) | RM | R | MAE (mm/d) | RMSE (mm/d) | RM | R |
| **$T_{max}$，$T_{min}$，SDun，Wspd** | | | | | | | | | | | | |
| $MLP_o$ | 0.406 | 0.527 | 1.011 | 0.963 | 0.562 | 0.737 | 0.978 | 0.921 | 0.944 | 1.287 | 0.918 | 0.780 |
| $XGBoost_o$ | 0.321 | 0.421 | 1.001 | 0.976 | 0.321 | 0.426 | 0.993 | 0.973 | 0.933 | 1.265 | 0.924 | 0.785 |
| $LightGBM_o$ | 0.360 | 0.472 | 1.001 | 0.969 | 0.361 | 0.473 | 0.995 | 0.966 | **0.935** | **1.261** | **0.923** | **0.787** |
| $CatBoost1_o$ | 0.392 | 0.507 | 1.002 | 0.965 | 0.401 | 0.519 | 0.987 | 0.960 | 0.942 | 1.264 | 0.916 | 0.787 |
| **$T_{max}$，$T_{min}$，SDun** | | | | | | | | | | | | |
| $MLP_o$ | 0.474 | 0.621 | 0.996 | 0.947 | 0.591 | 0.789 | 0.968 | 0.906 | 0.932 | 1.284 | 0.894 | 0.786 |
| $XGBoost_o$ | 0.457 | 0.600 | 0.998 | 0.950 | 0.459 | 0.605 | 0.986 | 0.945 | **0.923** | **1.267** | **0.894** | **0.791** |
| $LightGBM_o$ | 0.442 | 0.579 | 1.001 | 0.954 | 0.446 | 0.581 | 0.992 | 0.950 | 0.928 | 1.274 | 0.895 | 0.786 |
| $CatBoost1_o$ | 0.487 | 0.636 | 1.002 | 0.943 | 0.484 | 0.641 | 0.987 | 0.937 | 0.928 | 1.269 | 0.889 | 0.792 |
| **$T_{max}$，$T_{min}$，Wspd** | | | | | | | | | | | | |
| $MLP_o$ | 0.537 | 0.705 | 0.988 | 0.930 | 0.635 | 0.840 | 1.004 | 0.895 | 0.929 | 1.263 | 0.960 | 0.778 |
| $XGBoost_o$ | 0.464 | 0.612 | 0.999 | 0.948 | 0.450 | 0.603 | 1.007 | 0.946 | 0.928 | 1.263 | 0.989 | 0.779 |
| $LightGBM_o$ | 0.478 | 0.631 | 0.999 | 0.944 | 0.458 | 0.606 | 1.009 | 0.944 | 0.929 | 1.252 | 0.995 | 0.784 |
| $CatBoost1_o$ | 0.560 | 0.727 | 0.999 | 0.927 | 0.552 | 0.721 | 1.010 | 0.922 | **0.904** | **1.204** | **0.984** | **0.796** |
| **$T_{max}$，$T_{min}$** | | | | | | | | | | | | |
| $MLP_o$ | 0.635 | 0.840 | 1.039 | 0.911 | 0.728 | 0.970 | 1.047 | 0.878 | 0.922 | 1.268 | 0.994 | 0.785 |
| $XGBoost_o$ | 0.592 | 0.774 | 1.000 | 0.915 | 0.584 | 0.765 | 1.003 | 0.911 | 0.900 | 1.232 | 0.960 | 0.792 |
| $LightGBM_o$ | 0.600 | 0.783 | 1.000 | 0.914 | 0.582 | 0.767 | 1.001 | 0.909 | 0.918 | 1.254 | 0.953 | 0.786 |
| $CatBoost1_o$ | 0.623 | 0.808 | 1.000 | 0.907 | 0.613 | 0.806 | 1.002 | 0.900 | **0.886** | **1.208** | **0.956** | **0.798** |

Note: The statistical indicators of the best performing machine learning models with different input combinations for this climate zone are highlighted in blue, and the statistical indicators of the best performing machine learning models with the same input combination for this climate zone are highlighted in grey.

**Table 10**

Average statistics of the predictions with a lead time of 1-7 days by the $MLP_o$, $XGBoost_o$, $LightGBM_o$ and $CatBoost1_o$ models for the SHZ climate zone with different input combinations during training, validation and testing.

| Inputs/model | training | | | | validation | | | | testing | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | MAE (mm/d) | RMSE (mm/d) | RM | R | MAE (mm/d) | RMSE (mm/d) | RM | R | MAE (mm/d) | RMSE (mm/d) | RM | R |
| **$T_{max}$，$T_{min}$，SDun，Wspd** | | | | | | | | | | | | |
| $MLP_o$ | 0.290 | 0.380 | 0.993 | 0.971 | 0.293 | 0.383 | 0.959 | 0.969 | 0.890 | 1.208 | 1.051 | 0.695 |
| $XGBoost_o$ | 0.261 | 0.340 | 1.004 | 0.977 | 0.257 | 0.340 | 0.979 | 0.974 | 0.807 | 1.084 | 0.957 | 0.736 |
| $LightGBM_o$ | 0.277 | 0.364 | 1.004 | 0.974 | 0.276 | 0.360 | 0.976 | 0.971 | 0.804 | 1.080 | 1.024 | 0.738 |
| $CatBoost1_o$ | 0.338 | 0.439 | 1.003 | 0.962 | 0.337 | 0.435 | 0.985 | 0.957 | **0.786** | **1.044** | **0.986** | **0.750** |
| **$T_{max}$，$T_{min}$，SDun** | | | | | | | | | | | | |

https://doi.org/10.5194/hess-2023-158
Preprint. Discussion started: 31 July 2023
© Author(s) 2023. CC BY 4.0 License.
Hydrology and Earth System Sciences
Open Access
EGU
Discussions

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $MLP_o$ | 0.349 | 0.465 | 0.968 | 0.959 | 0.333 | 0.434 | 0.974 | 0.958 | 0.793 | 1.085 | 0.928 | 0.734 |
| $XGBoost_o$ | 0.339 | 0.449 | 0.986 | 0.961 | 0.316 | 0.418 | 0.995 | 0.961 | 0.776 | 1.069 | 0.986 | 0.743 |
| $LightGBM_o$ | 0.344 | 0.459 | 0.998 | 0.958 | 0.329 | 0.436 | 1.010 | 0.957 | 0.780 | 1.067 | 0.960 | 0.739 |
| $CatBoost1_o$ | 0.371 | 0.495 | 0.998 | 0.950 | 0.365 | 0.477 | 1.011 | 0.948 | **0.775** | **1.054** | **0.946** | **0.748** |
| **$T_{max}$，$T_{min}$，Wspd** | | | | | | | | | | | | |
| $MLP_o$ | 0.393 | 0.530 | 1.004 | 0.946 | 0.386 | 0.514 | 0.990 | 0.943 | 0.944 | 1.281 | 1.117 | 0.697 |
| $XGBoost_o$ | 0.376 | 0.497 | 1.002 | 0.950 | 0.351 | 0.467 | 0.991 | 0.950 | 0.824 | 1.118 | 1.047 | 0.736 |
| $LightGBM_o$ | 0.399 | 0.524 | 1.001 | 0.945 | 0.371 | 0.493 | 0.996 | 0.944 | 0.818 | 1.095 | 1.037 | 0.739 |
| $CatBoost1_o$ | 0.431 | 0.563 | 1.000 | 0.936 | 0.397 | 0.520 | 1.001 | 0.937 | **0.820** | **1.088** | **1.043** | **0.741** |
| **$T_{max}$，$T_{min}$** | | | | | | | | | | | | |
| $MLP_o$ | 0.449 | 0.601 | 1.013 | 0.927 | 0.441 | 0.585 | 1.049 | 0.925 | 0.806 | 1.116 | 1.012 | 0.732 |
| $XGBoost_o$ | 0.458 | 0.605 | 0.994 | 0.925 | 0.432 | 0.563 | 1.033 | 0.928 | 0.776 | 1.069 | 0.986 | 0.743 |
| $LightGBM_o$ | 0.472 | 0.618 | 0.994 | 0.923 | 0.444 | 0.575 | 1.030 | 0.926 | 0.772 | 1.059 | 0.991 | 0.746 |
| $CatBoost1_o$ | 0.473 | 0.621 | 0.994 | 0.921 | 0.446 | 0.579 | 1.030 | 0.924 | **0.772** | **1.059** | **0.993** | **0.748** |

Note: The statistical indicators of the best performing machine learning models with different input combinations for this climate zone are highlighted in blue, and the statistical indicators of the best performing machine learning models with the same input combination for this climate zone are highlighted in grey.

### 3.3. Performance comparison of five machine learning models trained and validated with 1-day ahead public weather forecast data for predicting $ET_o$ with various input combinations

The predicted daily $ET_o$ performance statistics of the five machine learning models trained and validated with daily public weather forecast data with a 1-day lead time from 2014-2018 and 2019, respectively, using various input combinations for the three climate zones, AR, SAR and SHZ, are shown in Tables 11-14. The predicted daily $ET_o$ performance varied significantly depending on the machine learning model type, input combination, and climate zone.

For the AR climate zone, the $ET_o$ prediction performance of $XGBoost_p$ and $LightGBM_p$ was the best in the training and validation periods with the C1, C2, and C3 input combinations, and $CatBoost1_p$ was the best for the C4 input combination. In the testing period, the $ET_o$ prediction performance of $MLP_p$, $XGBoost_p$, $LightGBM_p$, and $CatBoost1_p$ decreased across input combinations C4, C1, C2, and C3, C2, C1, C4, and C3, C1, C2, C3, and C4, and C1, C3, C2 and C4, respectively. $CatBoost1_p$ was the best performing model in terms of prediction performance (MAE=0.735 mm $d^{-1}$, RMSE=1.004 mm $d^{-1}$, and R=0.855) with the C3 input combination, and

https://doi.org/10.5194/hess-2023-158
Preprint. Discussion started: 31 July 2023
© Author(s) 2023. CC BY 4.0 License.

Hydrology and Earth System Sciences
Open Access
EGU
Discussions

$XGBoost_p$ had the best prediction performance (MAE range: 0.700-0.743 mm $d^{-1}$, RMSE range: 0.976-0.991 mm $d^{-1}$, and R range: 0.856-0.867) with the C1, C2, and C4 input combinations. $XGBoost_p$ with the C2 input combination was the best performing machine learning model in the AR climate zone testing period (MAE=0.703 mm $d^{-1}$, RMSE=0.976 mm $d^{-1}$, and R= 0.867).

For the SAR climate zone, in the training and validation periods, $XGBoost_p$ and $LightGBM_p$ were the best in terms of prediction performance with the C1, C2, and C3 input combinations, and $CatBoost1_p$ performed best with the C4 input combination. In the testing period, the $ET_o$ prediction performance of $MLP_p$, $XGBoost_p$, $LightGBM_p$, and $CatBoost1_p$ decreased across input combinations C4, C3, C2, and C1, C1, C2, C4, and C3, C2, C1, C4, and C3, and C3, C4, C1 and C2, respectively. $XGBoost_p$ had the best prediction performance (MAE range: 0.847-0.851 mm $d^{-1}$, RMSE range: 1.150-1.156 mm $d^{-1}$, and R range: 0.817-0.824) with both the C1 and C2 input combinations. $CatBoost1_p$ had the best prediction performance (MAE=0.860 mm $d^{-1}$, RMSE=1.177 mm $d^{-1}$, and R=0.813) with the C3 input combination. $MLP_p$ with the C4 input combination was the best performing machine learning model in the SAR climate zone testing period (MAE=0.850 mm $d^{-1}$, RMSE=1.148 mm $d^{-1}$, and R=0.818).

For the SHZ climate zone, in the training and validation periods, $XGBoost_p$ had the best prediction performance with the C2 and C3 input combinations, $LightGBM_p$ had the best prediction performance with the C1, C2 and C3 input combinations, and $CatBoost1_p$ had the best prediction performance with the C4 input combination. In the testing period, the $ET_o$ prediction performance of $MLP_p$, $XGBoost_p$, $LightGBM_p$, and $CatBoost1_p$ decreased across input combinations C4, C3, C2, and C1, C4, C1, C2, and C3, C1, C2, C4, and C3, and C3, C1, C4, and C2, respectively. The prediction performance of $XGBoost_p$ with the C1, C2, C3, and C4 input combinations (MAE range:

https://doi.org/10.5194/hess-2023-158
Preprint. Discussion started: 31 July 2023
© Author(s) 2023. CC BY 4.0 License.
Hydrology and Earth System Sciences
Open Access
Discussions

0.741-0.756 mm $d^{-1}$, RMSE range: 0.991-1.022 mm $d^{-1}$, and R range: 0.765-0.774) was the best, and $XGBoost_p$ with the C4 input combination was the best performing machine learning model in the SHZ climate zone testing period (MAE=0.756 mm $d^{-1}$, RMSE=0.991 mm $d^{-1}$, and R=0.774).

Table 14 shows the average statistics of the CatBoost2 model in predicting daily $ET_o$ 1-7 days ahead with different input combinations and the addition of wind scale (WS) and weather type (WT1 and WT2) category data. First, for the AR climate zone, the CatBoost2 model performed best with the C5 and C6 input combinations during the training and validation periods, and the performance decreased across input combinations C8, C5, C6, C7, and C9 during the testing period. The CatBoost2 model with the C8 input combination had the best performance in the testing period for the AR climate zone (MAE = 0.773 mm $d^{-1}$, RMSE = 1.059 mm $d^{-1}$, and R = 0.841). For the SAR climate zone, the CatBoost2 model in the training and validation periods had the best prediction performance with the C5 and C6 input combinations. In the testing period, the performance of the CatBoost2 model decreased across input combinations C8, C9, C7, C6, and C5. The CatBoost2 model with the C8 input combination had the best performance in the testing period for the SAR climate zone (MAE=0.904 mm $d^{-1}$, RMSE=1.241 mm $d^{-1}$, and R=0.798). For the SHZ climate zone, during the training and validation periods, the CatBoost2 model with the C5 and C6 input combinations had the best prediction performance, and during the testing period, the performance of the CatBoost2 model decreased across input combinations C8, C9, C6, C7, and C5. The CatBoost2 model with the C8 input combination had the best performance in the testing period for the SAR climate zone (MAE=0.793 mm $d^{-1}$, RMSE=1.057 mm $d^{-1}$, and R= 0.754). For the AR climate zone, $CatBoost1_p$ with the C2 input combination outperformed CatBoost2 with the C7 and C9 input combinations, and CatBoost2 with the C5, C6, and C8 input combinations outperformed

© Author(s) 2023. CC BY 4.0 License.


$CatBoost1_p$ with the C1 and C3 input combinations in the training and validation periods. In the testing period, $CatBoost1_p$ with the C1, C3, and C2 input combinations outperformed CatBoost2 with the C5, C6, C8, C7, and C9 input combinations. For the SAR climate region, $CatBoost1_p$ with the C2 input combination outperformed CatBoost2 with the C7 input combination, and CatBoost2 with the C5, C6, C8 and C9 input combinations outperformed $CatBoost1_p$ with the C1, C3 and C2 input combinations in both the training and validation periods. In the testing period, $CatBoost1_p$ with the C1, C3 and C2 input combinations outperformed CatBoost2 performance with the C5, C6, C8, C7 and C9 input combinations. For the SHZ climate zone, $CatBoost1_p$ with the C1 and C2 input combinations outperformed CatBoost2 with the C5, C7 and C9 input combinations, and CatBoost2 with the C6 and C8 input combinations outperformed $CatBoost1_p$ with the C1 and C3 input combinations in the training and validation periods. In the testing period, $CatBoost1_p$ with the C1, C3 and C2 input combinations outperformed CatBoost2 with the C5, C6, C8, C7, and C9 input combinations. These results show that although the CatBoost2 model outperformed the $CatBoost1_p$ model with some input combinations in the training and validation periods, the $CatBoost1_p$ model outperformed the CatBoost2 model with all input combinations in the testing period. Thus, adding category data such as wind scale (WS) and weather type (WT1 and WT2) directly to the input combinations of the CatBoost2 model did not improve the performance of the model in terms of predicting daily $ET_o$ in the testing period. This may be due to the poor stability of the CatBoost2 model and the poor performance of wind scale and weather type predictions in the public weather forecasts for the 1-7 days ahead period.

The optimal input combinations for four machine learning models ($MLP_o$, $XGBoost_o$, $LightGBM_o$, and $CatBoost1_o$) trained and validated with daily observed weather data and five

https://doi.org/10.5194/hess-2023-158
Preprint. Discussion started: 31 July 2023
© Author(s) 2023. CC BY 4.0 License.



machine learning models (MLPp, $XGBoost_p$, $LightGBM_p$, $CatBoost1_p$, and CatBoost2) trained and validated with 1-day ahead public weather forecast data for the three climate zones are shown in Table 15. In all climate zones except for SAR and SHZ, where the best input combination for CatBoost1p was C3 (8.33%), the best input combination for each of the eight machine learning models ($MLP_o$, $XGBoost_o$, $LightGBM_o$, $CatBoost1_o$, $MLP_p$, $XGBoost_p$, $LightGBM_p$, and $CatBoost1_p$) were either C1 (20.83%), C2 (29.17%), or C4 (41.67%). Consistent with previous findings (Yang et al., 2019a), the results indicate that the inclusion of SDun in the input combinations improves the performance of machine learning models in predicting daily $ET_o$ (Perera et al., 2014; Traore et al., 2016; Yang et al., 2016; Yang et al., 2019b; Yin et al., 2020; Zhou et al., 2020; Dong et al., 2021; Zhao et al., 2022), while the inclusion of Wspd leads to a decrease in the performance. Therefore, SDun and Wspd should be included in the input combinations of machine learning models cautiously. In addition, the model information obtained using the optimal input combination for each of the machine learning models ($MLP_p$, $XGBoost_p$, $LightGBM_p$, $CatBoost1_p$, and CatBoost2) trained and validated with 1-day ahead public weather forecast data is shown in Table 16. It is shown that the daily $ET_o$ performance predicted by MLP models with 2-3 hidden layers is better than that with 1 hidden layer (Luo et al., 2015; Traore et al., 2016; Ferreira et al., 2019).

**Table 11**

Mean statistics of 1-7 day lead time predictions using the $MLP_p$, $XGBoost_p$, $LightGBM_p$ and $CatBoost1_p$ models for the AR climate zone with the different input combinations during training, validation and testing.

| Inputs/model | training | | | | validation | | | | testing | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | MAE (mm/d) | RMSE (mm/d) | RM | R | MAE (mm/d) | RMSE (mm/d) | RM | R | MAE (mm/d) | RMSE (mm/d) | RM | R |
| **$T_{max}$, $T_{min}$, SDun, Wspd** | | | | | | | | | | | | |
| $MLP_p$ | 0.637 | 0.844 | 0.977 | 0.911 | 0.631 | 0.848 | 0.995 | 0.892 | 0.772 | 1.058 | 0.932 | 0.843 |
| $XGBoost_p$ | 0.410 | 0.559 | 0.982 | 0.963 | 0.404 | 0.553 | 0.990 | 0.953 | **0.700** | **0.976** | **0.961** | **0.865** |
| $LightGBM_p$ | 0.423 | 0.575 | 0.999 | 0.957 | 0.422 | 0.574 | 1.002 | 0.948 | 0.711 | 0.991 | 0.962 | 0.859 |

https://doi.org/10.5194/hess-2023-158
Preprint. Discussion started: 31 July 2023
© Author(s) 2023. CC BY 4.0 License.

Hydrology and Earth System Sciences
Open Access
Discussions

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CatBoost1$_p$ | 0.550 | 0.730 | 0.997 | 0.929 | 0.551 | 0.736 | 1.014 | 0.915 | 0.729 | 0.999 | 0.967 | 0.854 |
| **$T_{max}$，$T_{min}$，SDun** | | | | | | | | | | | | |
| MLP$_p$ | 0.646 | 0.854 | 1.005 | 0.903 | 0.676 | 0.885 | 1.056 | 0.883 | 0.787 | 1.050 | 1.046 | 0.843 |
| XGBoost$_p$ | 0.420 | 0.569 | 0.987 | 0.960 | 0.425 | 0.573 | 1.020 | 0.950 | **0.703** | **0.976** | **1.034** | **0.867** |
| LightGBM$_p$ | 0.434 | 0.593 | 0.994 | 0.954 | 0.444 | 0.606 | 1.028 | 0.944 | 0.716 | 0.996 | 1.040 | 0.863 |
| CatBoost1$_p$ | 0.572 | 0.766 | 0.993 | 0.922 | 0.588 | 0.781 | 1.036 | 0.906 | 0.748 | 1.006 | 1.036 | 0.854 |
| **$T_{max}$，$T_{min}$，Wspd** | | | | | | | | | | | | |
| MLP$_p$ | 0.725 | 0.959 | 1.059 | 0.889 | 0.756 | 1.004 | 1.081 | 0.861 | 0.794 | 1.081 | 1.032 | 0.849 |
| XGBoost$_p$ | 0.670 | 0.889 | 0.985 | 0.897 | 0.651 | 0.876 | 0.984 | 0.877 | 0.756 | 1.032 | 0.950 | 0.848 |
| LightGBM$_p$ | 0.663 | 0.879 | 0.998 | 0.897 | 0.668 | 0.883 | 1.008 | 0.875 | 0.743 | 1.009 | 0.973 | 0.851 |
| CatBoost1$_p$ | 0.686 | 0.909 | 0.997 | 0.889 | 0.706 | 0.931 | 1.016 | 0.861 | **0.735** | **1.004** | **0.982** | **0.855** |
| **$T_{max}$，$T_{min}$** | | | | | | | | | | | | |
| MLP$_p$ | 0.709 | 0.947 | 0.992 | 0.878 | 0.740 | 0.977 | 1.036 | 0.849 | 0.743 | 0.998 | 1.034 | 0.855 |
| XGBoost$_p$ | 0.731 | 0.966 | 0.950 | 0.882 | 0.710 | 0.939 | 0.986 | 0.857 | **0.743** | **0.991** | **0.995** | **0.856** |
| LightGBM$_p$ | 0.700 | 0.932 | 0.993 | 0.883 | 0.718 | 0.952 | 1.034 | 0.856 | 0.760 | 1.014 | 1.043 | 0.852 |
| CatBoost1$_p$ | 0.668 | 0.892 | 0.995 | 0.893 | 0.686 | 0.910 | 1.028 | 0.868 | 0.761 | 1.021 | 1.047 | 0.851 |

Note: The statistical indicators of the best performing machine learning models with different input combinations for this climate zone are highlighted in blue, and the statistical indicators of the best performing machine learning models for the same input combination for this climate zone are highlighted in grey.

**Table 12**

Mean statistics of 1-7 day lead time predictions using the MLP$_p$, XGBoost$_p$, LightGBM$_p$ and CatBoost1$_p$ models for the SAR climate zone with the different input combinations during training, validation and testing.

| Inputs/model | training | | | | validation | | | | testing | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | MAE (mm/d) | RMSE (mm/d) | RM | R | MAE (mm/d) | RMSE (mm/d) | RM | R | MAE (mm/d) | RMSE (mm/d) | RM | R |
| **$T_{max}$，$T_{min}$，SDun，Wspd** | | | | | | | | | | | | |
| MLP$_p$ | 0.641 | 0.850 | 1.014 | 0.900 | 0.639 | 0.860 | 1.012 | 0.887 | 0.910 | 1.245 | 0.909 | 0.801 |
| XGBoost$_p$ | 0.465 | 0.631 | 0.984 | 0.947 | 0.448 | 0.599 | 0.987 | 0.943 | **0.851** | **1.150** | **0.935** | **0.824** |
| LightGBM$_p$ | 0.468 | 0.639 | 0.998 | 0.945 | 0.453 | 0.613 | 1.004 | 0.940 | 0.866 | 1.182 | 0.945 | 0.816 |
| CatBoost1$_p$ | 0.583 | 0.778 | 0.998 | 0.914 | 0.575 | 0.769 | 1.009 | 0.904 | 0.876 | 1.191 | 0.938 | 0.810 |
| **$T_{max}$，$T_{min}$，SDun** | | | | | | | | | | | | |
| MLP$_p$ | 0.667 | 0.891 | 1.004 | 0.888 | 0.670 | 0.880 | 1.032 | 0.877 | 0.893 | 1.210 | 0.996 | 0.802 |
| XGBoost$_p$ | 0.467 | 0.636 | 0.991 | 0.944 | 0.459 | 0.617 | 1.007 | 0.939 | **0.847** | **1.156** | **0.976** | **0.817** |
| LightGBM$_p$ | 0.493 | 0.675 | 0.996 | 0.937 | 0.491 | 0.662 | 1.018 | 0.930 | 0.857 | 1.173 | 0.995 | 0.816 |
| CatBoost1$_p$ | 0.613 | 0.821 | 0.995 | 0.906 | 0.613 | 0.804 | 1.023 | 0.897 | 0.888 | 1.191 | 0.982 | 0.806 |
| **$T_{max}$，$T_{min}$，WSspd** | | | | | | | | | | | | |
| MLP$_p$ | 0.734 | 0.974 | 1.048 | 0.874 | 0.739 | 0.999 | 1.006 | 0.845 | 0.887 | 1.213 | 0.958 | 0.808 |
| XGBoost$_p$ | 0.673 | 0.900 | 1.001 | 0.885 | 0.656 | 0.897 | 0.996 | 0.867 | 0.894 | 1.216 | 0.921 | 0.806 |
| LightGBM$_p$ | 0.647 | 0.865 | 1.001 | 0.893 | 0.629 | 0.855 | 0.995 | 0.878 | 0.887 | 1.209 | 0.930 | 0.805 |
| CatBoost1$_p$ | 0.715 | 0.944 | 0.998 | 0.873 | 0.731 | 0.964 | 1.012 | 0.846 | **0.860** | **1.177** | **0.948** | **0.813** |
| **$T_{max}$，$T_{min}$** | | | | | | | | | | | | |
| MLP$_p$ | 0.746 | 1.003 | 1.018 | 0.857 | 0.761 | 1.010 | 1.031 | 0.835 | **0.850** | **1.148** | **0.995** | **0.818** |
| XGBoost$_p$ | 0.758 | 1.011 | 0.969 | 0.858 | 0.732 | 0.972 | 0.989 | 0.842 | 0.866 | 1.159 | 0.956 | 0.820 |
| LightGBM$_p$ | 0.749 | 0.996 | 0.997 | 0.858 | 0.740 | 0.984 | 1.016 | 0.839 | 0.877 | 1.174 | 0.979 | 0.811 |

© Author(s) 2023. CC BY 4.0 License.



| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $CatBoost1_p$ | 0.702 | 0.942 | 0.998 | 0.874 | 0.710 | 0.941 | 1.011 | 0.854 | 0.871 | 1.176 | 0.983 | 0.809 |

590 Note: The statistical indicators of the best performing machine learning models with different input combinations for this climate zone are
591 highlighted in blue, and the statistical indicators of the best performing machine learning models for the same input combination for this
592 climate zone are highlighted in grey.

593 **Table 13**

594 Mean statistics of 1-7 day lead time predictions using the $MLP_p$, $XGBoost_p$, $LightGBM_p$ and $CatBoost1_p$ models for the SHZ climate zone
595 with the different input combinations during training, validation and testing.

| Inputs/model | training | | | | validation | | | | testing | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | MAE (mm/d) | RMSE (mm/d) | RM | R | MAE (mm/d) | RMSE (mm/d) | RM | R | MAE (mm/d) | RMSE (mm/d) | RM | R |
| **$T_{max}$，$T_{min}$，SDun，Wspd** | | | | | | | | | | | | |
| $MLP_p$ | 0.511 | 0.675 | 1.033 | 0.905 | 0.498 | 0.680 | 1.069 | 0.895 | 0.778 | 1.066 | 0.976 | 0.761 |
| $XGBoost_p$ | 0.407 | 0.551 | 0.965 | 0.942 | 0.360 | 0.500 | 0.996 | 0.940 | **0.748** | **1.006** | **0.939** | **0.773** |
| $LightGBM_p$ | 0.359 | 0.492 | 0.995 | 0.950 | 0.331 | 0.468 | 1.020 | 0.946 | 0.742 | 1.016 | 0.961 | 0.770 |
| $CatBoost1_p$ | 0.399 | 0.536 | 0.997 | 0.938 | 0.377 | 0.521 | 1.018 | 0.930 | 0.756 | 1.027 | 0.956 | 0.769 |
| **$T_{max}$，$T_{min}$，SDun** | | | | | | | | | | | | |
| $MLP_p$ | 0.527 | 0.701 | 1.000 | 0.894 | 0.521 | 0.695 | 1.064 | 0.885 | 0.772 | 1.054 | 1.029 | 0.758 |
| $XGBoost_p$ | 0.410 | 0.560 | 0.962 | 0.939 | 0.370 | 0.518 | 1.011 | 0.933 | **0.741** | **1.010** | **0.996** | **0.772** |
| $LightGBM_p$ | 0.402 | 0.553 | 0.991 | 0.936 | 0.371 | 0.526 | 1.041 | 0.934 | 0.748 | 1.026 | 1.029 | 0.770 |
| $CatBoost1_p$ | 0.505 | 0.671 | 0.991 | 0.903 | 0.489 | 0.659 | 1.048 | 0.895 | 0.767 | 1.033 | 1.020 | 0.760 |
| **$T_{max}$，$T_{min}$，Wspd** | | | | | | | | | | | | |
| $MLP_p$ | 0.575 | 0.763 | 0.952 | 0.879 | 0.553 | 0.768 | 0.961 | 0.850 | 0.775 | 1.046 | 0.882 | 0.770 |
| $XGBoost_p$ | 0.557 | 0.737 | 0.995 | 0.882 | 0.535 | 0.736 | 1.019 | 0.860 | **0.753** | **1.022** | **0.962** | **0.765** |
| $LightGBM_p$ | 0.555 | 0.729 | 0.999 | 0.885 | 0.531 | 0.727 | 1.008 | 0.864 | 0.755 | 1.025 | 0.934 | 0.767 |
| $CatBoost1_p$ | 0.577 | 0.755 | 0.997 | 0.876 | 0.566 | 0.763 | 1.017 | 0.849 | 0.759 | 1.021 | 0.938 | 0.773 |
| **$T_{max}$，$T_{min}$** | | | | | | | | | | | | |
| $MLP_p$ | 0.590 | 0.781 | 0.999 | 0.867 | 0.610 | 0.803 | 1.046 | 0.836 | 0.745 | 1.005 | 1.008 | 0.773 |
| $XGBoost_p$ | 0.629 | 0.822 | 0.949 | 0.859 | 0.596 | 0.773 | 0.999 | 0.844 | **0.756** | **0.991** | **0.967** | **0.774** |
| $LightGBM_p$ | 0.582 | 0.771 | 0.994 | 0.870 | 0.582 | 0.770 | 1.038 | 0.849 | 0.755 | 1.013 | 1.011 | 0.768 |
| $CatBoost1_p$ | 0.534 | 0.712 | 0.995 | 0.890 | 0.537 | 0.711 | 1.027 | 0.871 | 0.763 | 1.031 | 1.013 | 0.762 |

596 Note: The statistical indicators of the best performing machine learning models with different input combinations for this climate zone are
597 highlighted in blue, and the statistical indicators of the best performing machine learning models for the same input combination for this
598 climate zone are highlighted in grey.

599 **Table 14**

600 Mean statistics of 1-7 day lead time predictions using the CatBoost2 model for the AR, SAR and SHZ climate zones with different input
601 combinations during training, validation and testing.

| Inputs/model | training | | | | validation | | | | testing | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | MAE (mm/d) | RMSE (mm/d) | RM | R | MAE (mm/d) | RMSE (mm/d) | RM | R | MAE (mm/d) | RMSE (mm/d) | RM | R |
| **$T_{max}$，$T_{min}$，WT1，WS** | | | | | | | | | | | | |
| AR/CatBoost2 | 0.524 | 0.695 | 1.001 | 0.936 | 0.487 | 0.661 | 0.993 | 0.931 | 0.781 | **1.069** | **0.951** | **0.837** |
| SAR/CatBoost2 | 0.502 | 0.671 | 1.007 | 0.936 | 0.464 | 0.625 | 0.989 | 0.937 | 0.929 | 1.280 | 0.942 | 0.783 |
| SHZ/CatBoost2 | 0.441 | 0.593 | 0.997 | 0.925 | 0.385 | 0.534 | 0.987 | 0.929 | 0.832 | 1.124 | 0.965 | 0.721 |
| **$T_{max}$，$T_{min}$，WT2，WS** | | | | | | | | | | | | |

https://doi.org/10.5194/hess-2023-158
Preprint. Discussion started: 31 July 2023
© Author(s) 2023. CC BY 4.0 License.



| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AR/CatBoost2 | 0.522 | 0.695 | 1.002 | 0.936 | 0.492 | 0.661 | 0.990 | 0.931 | 0.785 | 1.078 | 0.952 | 0.835 |
| SAR/CatBoost2 | 0.535 | 0.719 | 0.999 | 0.930 | 0.498 | 0.671 | 0.983 | 0.929 | 0.934 | 1.277 | 0.928 | 0.785 |
| SHZ/CatBoost2 | 0.382 | 0.510 | 1.000 | 0.944 | 0.333 | 0.461 | 0.993 | 0.946 | 0.800 | 1.083 | 0.940 | 0.741 |
| **$T_{max}$，$T_{min}$，WT1** | | | | | | | | | | | | |
| AR/CatBoost2 | 0.603 | 0.805 | 0.992 | 0.914 | 0.608 | 0.805 | 1.029 | 0.899 | 0.814 | 1.092 | 1.070 | 0.837 |
| SAR/CatBoost2 | 0.664 | 0.888 | 0.995 | 0.889 | 0.669 | 0.880 | 1.017 | 0.874 | 0.914 | **1.236** | **1.010** | **0.794** |
| SHZ/CatBoost2 | 0.495 | 0.668 | 0.993 | 0.904 | 0.491 | 0.654 | 1.038 | 0.894 | 0.802 | 1.098 | 1.042 | 0.743 |
| **$T_{max}$，$T_{min}$，WS** | | | | | | | | | | | | |
| AR/CatBoost2 | 0.564 | 0.754 | 1.004 | 0.925 | 0.520 | 0.705 | 0.987 | 0.921 | **0.773** | **1.059** | **0.933** | **0.841** |
| SAR/CatBoost2 | 0.594 | 0.795 | 1.002 | 0.911 | 0.544 | 0.738 | 0.979 | 0.912 | **0.904** | **1.241** | **0.909** | **0.798** |
| SHZ/CatBoost2 | 0.489 | 0.656 | 1.000 | 0.908 | 0.445 | 0.613 | 0.981 | 0.905 | **0.793** | **1.057** | **0.912** | **0.754** |
| **$T_{max}$，$T_{min}$，WT2** | | | | | | | | | | | | |
| AR/CatBoost2 | 0.561 | 0.751 | 0.996 | 0.925 | 0.567 | 0.748 | 1.023 | 0.913 | 0.822 | 1.104 | 1.076 | 0.835 |
| SAR/CatBoost2 | 0.624 | 0.837 | 0.998 | 0.903 | 0.636 | 0.833 | 1.022 | 0.890 | 0.915 | 1.233 | 1.006 | 0.791 |
| SHZ/CatBoost2 | 0.510 | 0.681 | 0.998 | 0.900 | 0.525 | 0.696 | 1.050 | 0.882 | **0.784** | **1.070** | **1.036** | **0.751** |

Note: 1. The prediction performance of the input combinations Tmax, Tmin, SDun, and WS (category data), Tmax, Tmin, SDun, and WS (numerical data) and Tmax, Tmin, SDun, and Wspd is consistent, and the prediction performance of the input combinations Tmax, Tmin, and WS (numerical data) and Tmax, Tmin, and Wspd is consistent. Therefore, they are not listed separately in Table 14.

2. The statistical indicators of the best performing CatBoost2 model for this climate region at different input combinations are highlighted in blue, and the statistical indicators of better performing CatBoost2 models for this climate region with different input combinations are highlighted in grey.

**Table 15**

Optimal input combinations of nine machine learning methods at nine sites in the three climate zones

| Climate zone | $MLP_o$ / $MLP_p$ | $XGBoost_o$ / $XGBoost_p$ | $LightGBM_o$ / $LightGBM_p$ | $CatBoost1_o$ / $CatBoost1_p$ | CatBoost2 |
|---|---|---|---|---|---|
| | Inputs/ Inputs | Inputs/ Inputs | Inputs/ Inputs | Inputs/ Inputs | Inputs |
| AR | C2 / C4 | C2 / C2 | C4 / C1 | C2 / C1 | C8 |
| SAR | C4 / C4 | C4 / C1 | C4 / C2 | C4 / C3 | C8 |
| SHZ | C2 / C4 | C2 / C4 | C4 / C1 | C1 / C3 | C8 |

**Table 16**

Optimal input combinations and model tuning information of five models (MLPp, XGBoostp, LightGBMp, CatBoost1p, and CatBoost2) at nine sites in the three climate zones.

| Models | Climate zone | Station | Inputs | Model information |
|---|---|---|---|---|
| **$MLP_p$** | AR | HN | C4 | learning_rate=0.0023, layer_size=66, hidden_layers=2, model structure: 2-66-66-1 |
| | | YC | C4 | learning_rate=0.0055, layer_size=93, hidden_layers=2, model structure: 2-93-93-1 |
| | | ZW | C4 | learning_rate=0.0035, layer_size=94, hidden_layers=3, model structure: 2-94-94-94-1 |
| | | ZN | C4 | learning_rate=0.0075, layer_size=73, hidden_layers=2, model structure: 2-73-73-1 |
| | SAR | YAC | C4 | learning_rate=0.0021, layer_size=79, hidden_layers=3, model structure: 2-79-79-79-1 |
| | | HY | C4 | learning_rate=0.0023, layer_size=66, hidden_layers=2, model structure: 2-66-66-1 |
| | | TX | C4 | learning_rate=0.0024, layer_size=88, hidden_layers=3, model structure: 2-88-88-88-1 |
| | SHZ | GY | C4 | learning_rate=0.0059, layer_size=100, hidden_layers=3, model structure: 2-100-100-100-1 |
| | | XJ | C4 | learning_rate=0.0017, layer_size=90, hidden_layers=3, model structure: 2-90-90-90-1 |
| **$XGBoost_p$** | AR | HN | C2 | colsample_bytree=0.98, eta=0.33, gamma=1.06, max_depth=14, min_child_weight=8, n_estimators=132, reg_alpha=2.65, reg_lambda=14.6 |
| | | YC | C1 | colsample_bytree=0.64, eta=0.16, gamma=1.00, max_depth=9, min_child_weight=7, |

https://doi.org/10.5194/hess-2023-158
Preprint. Discussion started: 31 July 2023
© Author(s) 2023. CC BY 4.0 License.
Hydrology and Earth System Sciences
Open Access
EGU
Discussions

| | | | | |
|---|---|---|---|---|
| | | | | n_estimators=279, reg_alpha=1.23, reg_lambda=1.97 |
| | | ZW | C2 | colsample_bytree=0.85, eta=0.06, gamma=1.60, max_depth=15, min_child_weight=4, n_estimators=52, reg_alpha=0.03, reg_lambda=0.71 |
| | | ZN | C1 | colsample_bytree=0.51, eta=0.17, gamma=1.00, max_depth=5, min_child_weight=0, n_estimators=267, reg_alpha=4.00, reg_lambda=0.52 |
| | SAR | YAC | C2 | colsample_bytree=0.85, eta=0.17, gamma=1.02, max_depth=10, min_child_weight=4, n_estimators=277, reg_alpha=0.04, reg_lambda=8.55 |
| | | HY | C1 | colsample_bytree=0.58, eta=0.02, gamma=1.48, max_depth=12, min_child_weight=4, n_estimators=278, reg_alpha=0.55, reg_lambda=18.26 |
| | | TX | C4 | colsample_bytree=0.62, eta=0.04, gamma=4.21, max_depth=17, min_child_weight=1, n_estimators=94, reg_alpha=22.63, reg_lambda=20.36 |
| | SHZ | GY | C4 | colsample_bytree=0.64, eta=0.04, gamma=1.90, max_depth=17, min_child_weight=9, n_estimators=82, reg_alpha=22.99, reg_lambda=0.62 |
| | | XJ | C1 | colsample_bytree=0.64, eta=0.11, gamma=1.51, max_depth=4, min_child_weight=0, n_estimators=97, reg_alpha=0.02, reg_lambda=0.64 |
| **LightGBM$_p$** | AR | HN | C2 | colsample_bytree=0.99, lr=0.14, max_depth=8, min_child_weight=1.94, n_estimators=57, min_data_in_leaf=10, num_leaves=75, reg_alpha=1.26, reg_lambda=54.38, subsample=0.44 |
| | | YC | C1 | colsample_bytree=0.39, lr=0.41, max_depth=10,min_child_weight=38.46, n_estimators=270, min_data_in_leaf=54, num_leaves=43, reg_alpha=1.75, reg_lambda=12.02, subsample=0.33 |
| | | ZW | C1 | colsample_bytree=0.78, lr=0.06, max_depth=4, min_child_weight=23.26, n_estimators=190, min_data_in_leaf=22, num_leaves=157, reg_alpha=3.57, reg_lambda=3.52, subsample=0.08 |
| | | ZN | C1 | colsample_bytree=0.99, lr=0.23, max_depth=5, min_child_weight=27.40, n_estimators=209, min_data_in_leaf=53, num_leaves=150, reg_alpha=4.87, reg_lambda=21.22, subsample=0.26 |
| | SAR | YAC | C2 | colsample_bytree=0.86, lr=0.04, max_depth=8, min_child_weight=4.72, n_estimators=286, min_data_in_leaf=2, num_leaves=145, reg_alpha=0.06, reg_lambda=42.54, subsample=0.91 |
| | | HY | C1 | colsample_bytree=0.81, lr=0.22, max_depth=10, min_child_weight=20.85, n_estimators=54, min_data_in_leaf=26, num_leaves=127, reg_alpha=3.23, reg_lambda=58.77, subsample=0.29 |
| | | TX | C1 | colsample_bytree=0.78, lr=0.02, max_depth=9, min_child_weight=17.27, n_estimators=201, min_data_in_leaf=5, num_leaves=110, reg_alpha=0.02, reg_lambda=7.749, subsample=0.69 |
| | SHZ | GY | C1 | colsample_bytree=0.99, lr=0.57, max_depth=10, min_child_weight=27.72, n_estimators=252, min_data_in_leaf=30, num_leaves=72, reg_alpha=5.49, reg_lambda=55.88, subsample=0.14 |
| | | XJ | C1 | colsample_bytree=0.65, lr=0.02, max_depth=9, min_child_weight=0.16, n_estimators=166, min_data_in_leaf=5, num_leaves=109, reg_alpha=1.05, reg_lambda=6.31, subsample=0.65 |
| **CatBoost1$_p$** | AR | HN | C2 | lr=0.01, l2_leaf_reg=1.96, depth=3, boosting_type='Plain' |
| | | YC | C1 | lr=0.01, l2_leaf_reg=4.01, depth=3, boosting_type='Plain' |
| | | ZW | C1 | lr=0.01, l2_leaf_reg=1.94, depth=7, boosting_type='Ordered' |
| | | ZN | C3 | lr=0.02, l2_leaf_reg=7.47, depth=1, boosting_type='Plain', rs=1.88, od_pval=0.009 |
| | SAR | YAC | C2 | lr=0.01, l2_leaf_reg=4.16, depth=3, boosting_type='Plain' |
| | | HY | C3 | lr=0.09, l2_leaf_reg=6.68, depth=1, boosting_type='Ordered', rs=3.77, od_pval=0.004 |
| | | TX | C4 | lr=0.02, l2_leaf_reg=13.43, depth=1, boosting_type='Ordered', rs=1.88, od_pval=0.003 |
| | SHZ | GY | C4 | lr=0.07, l2_leaf_reg=16.70, depth=2, boosting_type='Ordered', rs=7.51, od_pval=0.003 |
| | | XJ | C3 | lr=0.03, l2_leaf_reg=10.48, depth=1, boosting_type='Ordered', rs=3.51, od_pval=0.002 |
| **CatBoost2** | AR | HN | C5 | lr=0.03, l2_leaf_reg=6.07, depth=3, boosting_type='Plain', max_ctr_complexity=6 |
| | | YC | C5 | lr=0.03, l2_leaf_reg=6.39, depth=8, boosting_type='Plain', max_ctr_complexity=7 |

https://doi.org/10.5194/hess-2023-158
Preprint. Discussion started: 31 July 2023
© Author(s) 2023. CC BY 4.0 License.
Hydrology and Earth System Sciences
Open Access
Discussions

| | | | |
|---|---|---|---|
| | ZW | C8 | lr=0.06, l2_leaf_reg=2.08, depth=7, boosting_type='Ordered', max_ctr_complexity=4 |
| | ZN | C8 | lr=0.07, l2_leaf_reg=4.26, depth=5, boosting_type='Plain', max_ctr_complexity=5 |
| SAR | YAC | C7 | lr=0.01, l2_leaf_reg=7.17, depth=3, boosting_type='Ordered', max_ctr_complexity=0 |
| | HY | C8 | lr=0.09, l2_leaf_reg=5.33, depth=8, boosting_type='Ordered', max_ctr_complexity=6 |
| | TX | C9 | lr=0.01, l2_leaf_reg=7.17, depth=3, boosting_type='Ordered', max_ctr_complexity=0 |
| SHZ | GY | C9 | lr=0.01, l2_leaf_reg=7.17, depth=3, boosting_type='Ordered', max_ctr_complexity=0 |
| | XJ | C8 | lr=0.06, l2_leaf_reg=2.56, depth=10, boosting_type='Ordered', max_ctr_complexity=7 |

**Note:** lr: learning_rate and rs: random_strength.

### 3.4. Performance evaluation of $ET_o$ predicted by models based on public weather forecasts

#### 3.4.1. Performance of daily $ET_o$ predicted by nine models

The performance metrics of four models developed based on daily observed meteorological data and five models developed based on public weather forecast data with a 1-day lead time to predict daily ETo with a 1-7 day lead time for the three climate zones are shown in Figures 4 and 5. First, the daily $ET_o$ prediction performance of the nine models for the three climate zones, AR, SAR, and SHZ, decreased with increasing lead time, which is due to the decrease in forecast performance of public weather forecasting variables with increasing lead time, which is consistent with previous studies (Perera et al., 2014; Luo et al., 2014 and 2015; Traore et al., 2016; Yang et al., 2016, 2019a, 2019b; Traore et al., 2017; Li et al., 2018; Yin et al., 2020). In addition, for all three climate zones, the four models developed based on public weather forecast data with a 1-day lead time to predict daily $ET_o$ 1-7 days ahead outperformed the four models developed based on daily observed meteorological data with corresponding input combinations (except for the 1-day ahead prediction performance of $MLP_p$, $XGBoost_p$, and $CatBoost1_p$ for the SHZ climate zone). Second, the RM values of the nine models for the AR climate zone varied in the range of 0.92-1.07. Three models, $MLP_p$, $XGBoost_p$ and $LightGBM_o$, slightly overestimated (2.90%-6.58%) the daily $ET_o$, while $LightGBM_p$, $CatBoost1_p$, CatBoost2, $MLP_o$, $XGBoost_o$ and $CatBoost1_o$ slightly underestimated (2.69%-7.71%) the daily $ET_o$. The RM values for the nine models varied in the range of 0.90-1.00

© Author(s) 2023. CC BY 4.0 License.


for the SAR climate zone, and all models slightly underestimated (0.18%-9.56%) the daily $ET_o$ (except $LightGBM_p$, which slightly overestimated (0.12%) the daily $ET_o$ 6-days ahead, and $MLP_o$, which slightly overestimated (0.02%) the daily $ET_o$ 7-days ahead. The RM values for the 9 models varied in the range of 0.90-1.02 for the SHZ climate zone. $MLP_p$ slightly overestimated (0.13%-1.64%) the daily $ET_o$, and all other models slightly underestimated (0.02%-9.94%) the daily $ET_o$.

Table 17 shows the 1-7 day ahead $ET_o$ prediction performance comparison utilizing the best input combinations in the testing period for the four models developed based on daily observed meteorological data and the five models developed based on public weather forecast data with a 1-day lead time for the three climate zones. Overall, for all three climate zones, the four models developed based on 1-day ahead public weather forecast data generally outperformed the four models developed based on daily observed meteorological data with corresponding input combinations for all metrics. In addition, the prediction performance of all models exhibited a decrease in the following order for the three climate zones: AR, SHZ, and SAR. This result is mainly because the prediction performance of models developed based on public weather forecasts for the AR climate zone is better than that for the SAR and SHZ climate zones. For the AR climate zone, the mean MAE and RMSE ranges of the four models (MLP, XGBoost, LightGBM, and CatBoost1) were 0.770-0.805 mm $d^{-1}$ and 1.042-1.081 mm $d^{-1}$ (performance of models trained and validated based on daily observed meteorological data) to 0.703-0.743 mm $d^{-1}$ and 0.976-0.999 mm $d^{-1}$ (performance of models trained and validated based on public weather forecast data with a 1-day lead time), respectively, a decrease of 5.32%-11.67% and 4.13%-7.68%, respectively. The mean R value range increased from 0.837-0.844 to 0.854-0.867, an improvement of 1.31%-3.46%. For the SAR climate zone, the mean MAE and RMSE ranges for the four models (MLP, XGBoost,

https://doi.org/10.5194/hess-2023-158
Preprint. Discussion started: 31 July 2023
© Author(s) 2023. CC BY 4.0 License.



LightGBM, and CatBoost1) decreased from 0.886-0.922 mm $d^{-1}$ and 1.208-1.268 mm $d^{-1}$ (performance of the models trained and validated based on daily observed meteorological data) to 0.850-0.860 mm $d^{-1}$ and 1.148-1.177 mm $d^{-1}$ (performance of models trained and validated based on public weather forecast data with 1-day lead time), respectively, a reduction of 2.93%-7.81% and 2.57%-9.46%, respectively. The mean R range increased from 0.785-0.798 to 0.813-0.824, an improvement of 1.88%-4.20%. For the SHZ climate zone, the mean MAE and RMSE ranges for the four models (MLP, XGBoost, LightGBM, and CatBoost1) decreased from 0.772-0.793 mm $d^{-1}$ and 1.044-1.085 mm $d^{-1}$ (performance of models trained and validated based on daily observed meteorological data) to 0.742-0.759 mm $d^{-1}$ and 0.991-1.021 mm $d^{-1}$ (performance of models trained and validated based on 1-day ahead public weather forecast data), respectively, a reduction of 3.89%-6.05% and 2.20%-7.37%, respectively. The mean R ranged increased from 0.734-0.750 to 0.770-0.774, an improvement of 3.07%-5.31%. Finally, when considering all metrics, the top three models in the AR climate zone were $XGBoost_p$, $LightGBM_p$, and $MLP_p$, while $MLP_o$ was the worst performing model; the top three models in the SAR climate zone were $MLP_p$, $XGBoost_p$, and $LightGBM_p$, while $MLP_o$ was the worst performing model; and the top three models in the SHZ climate zone were $XGBoost_p$, $MLP_p$ and $LightGBM_p$, while $MLP_o$ was the worst performing model.

https://doi.org/10.5194/hess-2023-158
Preprint. Discussion started: 31 July 2023
© Author(s) 2023. CC BY 4.0 License.

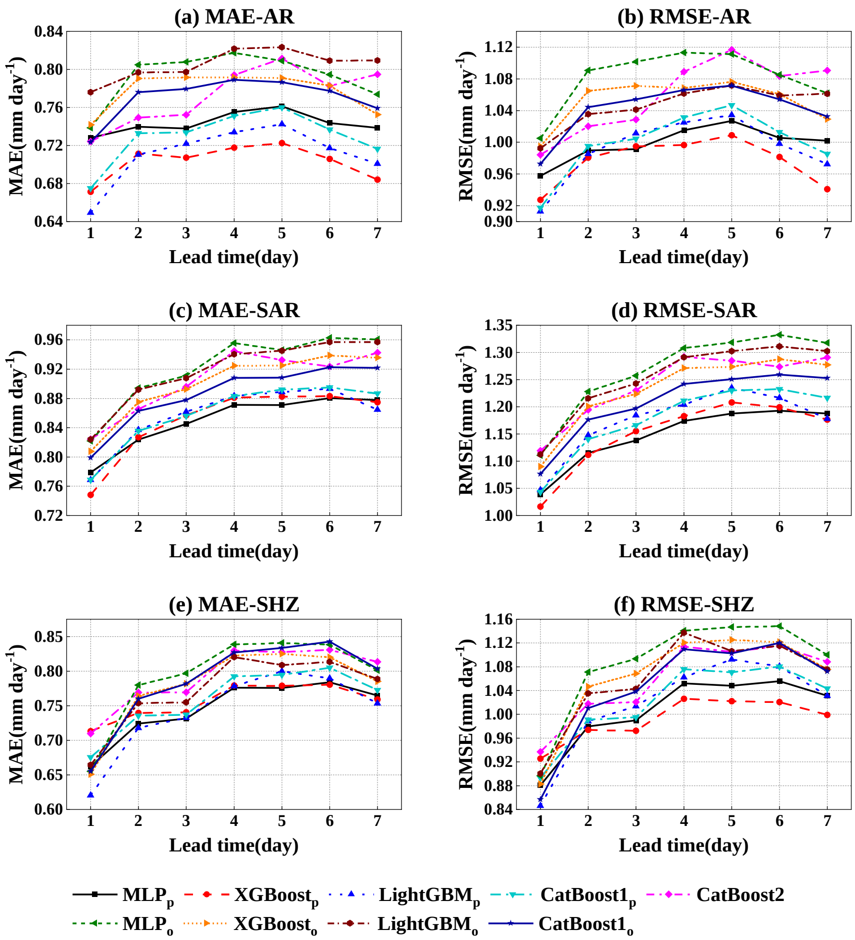




**Figure. 4.** The MAE and RMSE statistics for predicting $ET_o$ with a lead time of 1-7 days using four models trained and validated based on daily observed meteorological data and five models trained and validated based on 1-day ahead public weather forecast data for the three climate zones.

Hydrology and Earth System Sciences
Discussions
Open Access

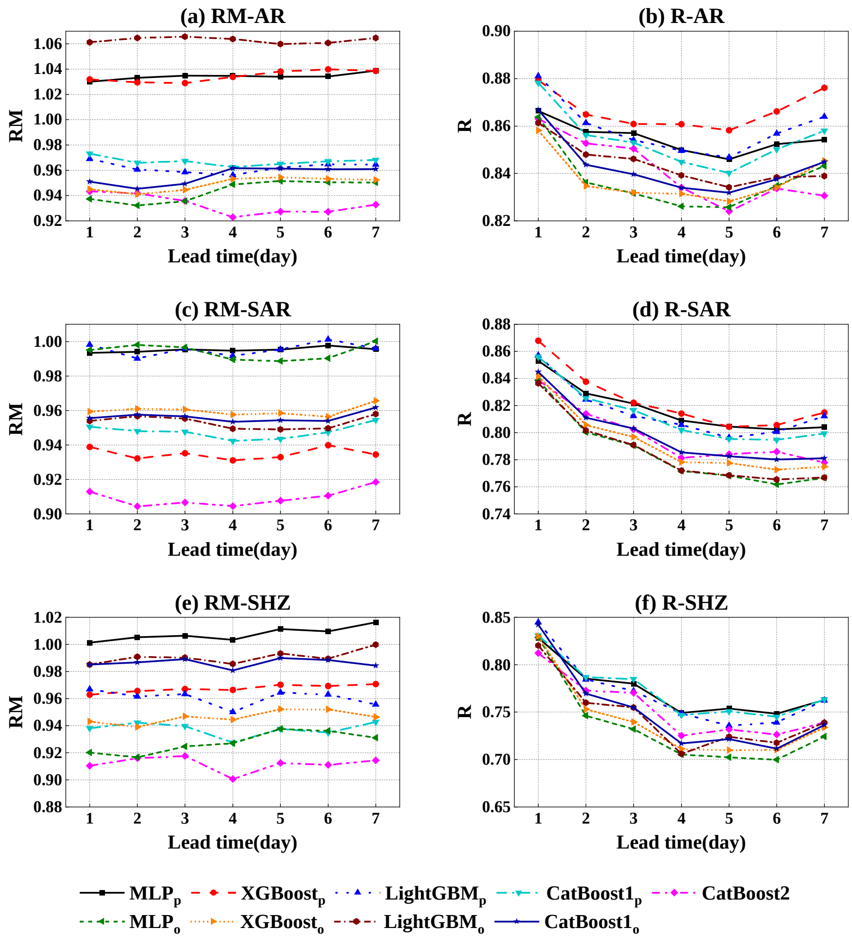



**Figure. 5.** The RM and R statistics for predicting ETo with a lead time of 1-7 days using four models trained and validated based on daily observed meteorological data and five models trained and validated based on 1-day ahead public weather forecast data for the three climate zones.

**Table 17**

Mean statistics of the performance indicators for ETo prediction with a lead time of 1–7 days using four models trained and validated based on daily observed meteorological data and five models trained and validated based on 1-day ahead public weather forecast data for three climate zones.

| 方法 | AR | | | | SAR | | | | SHZ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | MAE (mm/d) | RMSE (mm/d) | RM | R | MAE (mm/d) | RMSE (mm/d) | RM | R | MAE (mm/d) | RMSE (mm/d) | RM | R |
| MLP$_o$ | 0.792 | 1.081 | 0.944 | 0.837 | 0.922 | 1.268 | 0.994 | 0.785 | 0.793 | 1.085 | 0.928 | 0.734 |
| MLP$_p$ | 0.743 | 0.998 | 1.034 | 0.855 | 0.850 | 1.148 | 0.995 | 0.818 | 0.745 | 1.005 | 1.008 | 0.773 |
| XGBoost$_o$ | 0.777 | 1.052 | 0.949 | 0.838 | 0.900 | 1.232 | 0.960 | 0.792 | 0.776 | 1.069 | 0.986 | 0.743 |
| XGBoost$_p$ | 0.703 | 0.976 | 1.034 | 0.867 | 0.851 | 1.150 | 0.935 | 0.824 | 0.756 | 0.991 | 0.967 | 0.774 |
| LightGBM$_o$ | 0.805 | 1.046 | 1.063 | 0.844 | 0.918 | 1.254 | 0.953 | 0.786 | 0.772 | 1.059 | 0.991 | 0.746 |

https://doi.org/10.5194/hess-2023-158
Preprint. Discussion started: 31 July 2023
© Author(s) 2023. CC BY 4.0 License.
Hydrology and Earth System Sciences
Open Access
Discussions

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\text{LightGBM}_p$ | **0.711** | **0.991** | **0.962** | **0.859** | 0.857 | 1.173 | 0.995 | 0.816 | 0.742 | 1.016 | 0.961 | 0.770 |
| $\text{CatBoost1}_o$ | 0.770 | 1.042 | 0.956 | 0.843 | 0.886 | 1.208 | 0.956 | 0.798 | 0.786 | 1.044 | 0.986 | 0.750 |
| $\text{CatBoost1}_p$ | 0.729 | 0.999 | 0.967 | 0.854 | 0.860 | 1.177 | 0.948 | 0.813 | 0.759 | 1.021 | 0.938 | 0.773 |
| CatBoost2 | 0.773 | 1.059 | 0.933 | 0.841 | 0.904 | 1.241 | 0.909 | 0.798 | 0.793 | 1.057 | 0.912 | 0.754 |

Note: The statistical indicators of the best performance model for each climate zone are highlighted in blue, and the statistical indicators of the better performance model are highlighted in grey.

3.4.2. Seasonality of the performance of the five models developed based on public weather forecast data with a 1-day lead time to predict daily $ET_o$

Irrigated areas in the studied sites have different field irrigation seasons according to crop type and location (Perera et al., 2014). In addition, the sensitivity of weather variables varies with season and the microclimate of the studied site location (Vanella et al., 2020). Therefore, there is a strong need to evaluate the seasonality of the $ET_o$ forecast performance. Tables 18-20 show the mean statistics of the performance indicators for the 1-7 day lead time $ET_o$ forecasts using the five models developed based on public weather forecast data with a 1-day lead time for the three climatic zones for all seasons of 2020-2021. First, during all four seasons, the daily $ET_o$ prediction performance of all five models was better in the AR climate zone than in the SAR and SHZ climate zones (excluding CatBoost2 in the spring and MLP in the fall in the SHZ climate zone, and all five models in the winter in the SHZ climate zone). This is mainly because the prediction performance of models using the public weather forecast variables in the AR climate zone outperform those in the SAR and SHZ climate zones overall.

Second, the order of the seasonal MAE and RMSE values for daily $ET_o$ prediction using the five models in the AR climate zone are as follows: winter, spring, fall, and summer. In addition, the seasonal MAE and RMSE values during winter are lower than the annual average, except for the MLPp and LightGBMp models for station TX in the SAR climate zone and the MLPp and XGBoost models for station XJ in the SHZ climate zone. The order of the seasonal MAE and RMSE values

https://doi.org/10.5194/hess-2023-158
Preprint. Discussion started: 31 July 2023
© Author(s) 2023. CC BY 4.0 License.

Hydrology and Earth System Sciences
Discussions

using the other three models in the SAR and SHZ climate zones are as follows: winter, fall, spring, and summer. The seasonal MAE and RMSE values during winter and fall in the SAR climate zone are lower than the annual average, and the seasonal MAE and RMSE values during winter in the SHZ climate zone are lower than the annual average. These results are consistent with results in previous studies (Perera et al., 2014; Yang et al., 2016; Fan et al., 2021b). The seasonal R values of $ET_o$ for all model prediction days in all three climate zones were lower than the annual average, with the smallest R values observed during summer. The maximum seasonal R values in the AR climate zone occurred in spring (XGBoostp, LightGBMp, and CatBoost1p) and fall (MLPp and CatBoost2), those in the SAR climate zone occurred in spring (XGBoostp and LightGBMp) and fall (MLPp, CatBoost1p, and CatBoost2), and those for the SHZ climate zone occurred in spring (XGBoostp) and fall (MLPp, LightGBMp, CatBoost1p, and CatBoost2). The seasonal RM values of daily $ET_o$ prediction using the five machine learning models in the three climate zones were less than 1 in spring and summer and greater than 1 in fall and winter, indicating that daily $ET_o$ was underestimated in spring and summer (except for XGBoostp in summer in the AR climate zone, which was overestimated by 2.93%) and overestimated in fall and winter (except for CatBoost1p in winter in the SHZ climate zone and CatBoost2, which were underestimated by 7.75% and 2.05%, respectively).

Finally, when considering all the metrics, for the AR climate zone, XGBoostp, MLPp, CatBoost2, and LightGBMp showed the best performance in predicting 1-7 day ahead seasonal ET values during spring, summer, fall, and winter, respectively, with MLPp (spring), XGBoostp (summer), LightGBMp (fall), and CatBoost2 (winter) being the second best performers. For the SAR climate zone, LightGBMp, MLPp, CatBoost1p and CatBoost2 showed the best performance in predicting

https://doi.org/10.5194/hess-2023-158
Preprint. Discussion started: 31 July 2023
© Author(s) 2023. CC BY 4.0 License.



1-7 day ahead seasonal ET values during spring, summer, fall and winter, respectively, with MLPp (spring), XGBoostp (summer and fall) and CatBoost1p (winter) being the next best performers. For the SHZ climate zone, XGBoostp, CatBoost1p, and LightGBMp showed the best performance in predicting 1-7 day ahead seasonal ET values during spring, summer, fall, and winter, respectively, with MLPp (spring, summer, and winter) and XGBoostp (fall) being the next best performers.

**Table 18**

The mean statistics of the performance indicators for 1-7 days lead time $ET_o$ prediction using 5 methods at 4 stations in the AR climate region during the four seasons of 2020-2021.

| Stations/Methods | Spring | | | | Summer | | | | Fall | | | | Winter | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | MAE (°C) | RMSE (°C) | RM | R | MAE (°C) | RMSE (°C) | RM | R | MAE (°C) | RMSE (°C) | RM | R | MAE (°C) | RMSE (°C) | RM | R |
| **AR/Huinong** | | | | | | | | | | | | | | | | |
| MLP$_p$ | 0.717 | 0.933 | 0.879 | 0.744 | 0.980 | 1.225 | 0.945 | 0.092 | 0.885 | 1.117 | 1.185 | 0.726 | 0.406 | 0.500 | 1.260 | 0.490 |
| XGBoost$_p$ | 0.675 | 0.894 | 0.952 | 0.739 | 0.955 | 1.209 | 1.013 | 0.125 | 0.819 | 1.086 | 1.061 | 0.678 | 0.386 | 0.483 | 1.113 | 0.353 |
| LightGBM$_p$ | 0.734 | 0.987 | 0.859 | 0.728 | 1.004 | 1.270 | 0.949 | 0.130 | 0.819 | 1.067 | 0.965 | 0.670 | 0.347 | 0.483 | 1.018 | 0.367 |
| CatBoost1$_p$ | 0.747 | 0.972 | 0.853 | 0.742 | 1.023 | 1.258 | 0.913 | 0.110 | 0.836 | 1.071 | 1.125 | 0.695 | 0.430 | 0.516 | 1.292 | 0.448 |
| CatBoost2 | 0.914 | 1.164 | 0.771 | 0.696 | 1.184 | 1.475 | 0.894 | 0.114 | 0.787 | 1.019 | 1.008 | 0.693 | 0.354 | 0.479 | 1.023 | 0.400 |
| **AR/Yinchuan** | | | | | | | | | | | | | | | | |
| MLP$_p$ | 0.683 | 0.904 | 0.874 | 0.701 | 0.841 | 1.081 | 0.982 | 0.199 | 0.815 | 1.043 | 1.234 | 0.773 | 0.347 | 0.443 | 1.295 | 0.517 |
| XGBoost$_p$ | 0.613 | 0.837 | 0.942 | 0.709 | 0.852 | 1.111 | 1.027 | 0.186 | 0.728 | 0.976 | 1.135 | 0.736 | 0.308 | 0.393 | 1.208 | 0.473 |
| LightGBM$_p$ | 0.671 | 0.909 | 0.894 | 0.682 | 0.868 | 1.136 | 0.989 | 0.200 | 0.698 | 0.933 | 1.039 | 0.721 | 0.255 | 0.397 | 1.040 | 0.418 |
| CatBoost1$_p$ | 0.737 | 0.966 | 0.817 | 0.708 | 0.875 | 1.127 | 0.965 | 0.195 | 0.723 | 0.950 | 1.111 | 0.723 | 0.299 | 0.396 | 1.201 | 0.507 |
| CatBoost2 | 0.812 | 1.055 | 0.799 | 0.658 | 0.954 | 1.210 | 0.935 | 0.190 | 0.708 | 0.929 | 1.100 | 0.739 | 0.255 | 0.386 | 1.066 | 0.489 |
| **AR/Zhongwei** | | | | | | | | | | | | | | | | |
| MLP$_p$ | 0.862 | 1.108 | 0.846 | 0.655 | 1.070 | 1.349 | 1.002 | 0.078 | 0.882 | 1.143 | 1.232 | 0.715 | 0.373 | 0.491 | 1.234 | 0.471 |
| XGBoost$_p$ | 0.789 | 1.037 | 0.907 | 0.664 | 1.082 | 1.369 | 1.008 | 0.101 | 0.780 | 1.050 | 1.098 | 0.657 | 0.341 | 0.466 | 1.151 | 0.402 |
| LightGBM$_p$ | 0.883 | 1.147 | 0.817 | 0.660 | 1.093 | 1.378 | 0.970 | 0.123 | 0.774 | 1.033 | 1.040 | 0.639 | 0.331 | 0.445 | 1.149 | 0.444 |
| CatBoost1$_p$ | 0.890 | 1.146 | 0.815 | 0.671 | 1.089 | 1.372 | 0.966 | 0.122 | 0.781 | 1.052 | 1.050 | 0.639 | 0.322 | 0.436 | 1.097 | 0.433 |
| CatBoost2 | 1.016 | 1.277 | 0.749 | 0.657 | 1.158 | 1.421 | 0.936 | 0.091 | 0.777 | 1.020 | 1.114 | 0.693 | 0.331 | 0.455 | 1.097 | 0.425 |
| **AR/Zhongning** | | | | | | | | | | | | | | | | |
| MLP$_p$ | 0.702 | 0.912 | 0.885 | 0.681 | 1.026 | 1.318 | 1.015 | 0.145 | 0.885 | 1.145 | 1.226 | 0.716 | 0.394 | 0.495 | 1.307 | 0.539 |
| XGBoost$_p$ | 0.672 | 0.885 | 0.974 | 0.673 | 1.119 | 1.416 | 1.069 | 0.162 | 0.799 | 1.082 | 1.118 | 0.673 | 0.328 | 0.446 | 1.187 | 0.475 |
| LightGBM$_p$ | 0.743 | 0.971 | 0.864 | 0.662 | 1.090 | 1.373 | 1.001 | 0.170 | 0.762 | 1.025 | 1.013 | 0.661 | 0.279 | 0.402 | 1.075 | 0.507 |
| CatBoost1$_p$ | 0.811 | 1.036 | 0.805 | 0.679 | 1.087 | 1.378 | 0.984 | 0.165 | 0.746 | 1.014 | 1.036 | 0.668 | 0.291 | 0.391 | 1.113 | 0.519 |
| CatBoost2 | 0.879 | 1.110 | 0.777 | 0.665 | 1.197 | 1.478 | 0.948 | 0.174 | 0.727 | 0.974 | 1.069 | 0.696 | 0.287 | 0.408 | 1.036 | 0.524 |
| **AR/Average** | | | | | | | | | | | | | | | | |
| MLP$_p$ | 0.741 | 0.964 | 0.871 | 0.695 | 0.979 | 1.243 | 0.986 | 0.129 | 0.867 | 1.112 | 1.219 | 0.733 | 0.380 | 0.482 | 1.274 | 0.504 |
| XGBoost$_p$ | 0.687 | 0.913 | 0.944 | 0.696 | 1.002 | 1.276 | 1.029 | 0.144 | 0.782 | 1.049 | 1.103 | 0.686 | 0.341 | 0.447 | 1.165 | 0.426 |
| LightGBM$_p$ | 0.758 | 1.004 | 0.859 | 0.683 | 1.014 | 1.289 | 0.977 | 0.156 | 0.763 | 1.015 | 1.014 | 0.673 | 0.303 | 0.432 | 1.071 | 0.434 |

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CatBoost1$_p$ | 0.796 | 1.030 | 0.823 | 0.700 | 1.019 | 1.284 | 0.957 | 0.148 | 0.772 | 1.022 | 1.081 | 0.681 | 0.336 | 0.435 | 1.176 | 0.477 |
| CatBoost2 | 0.905 | 1.152 | 0.774 | 0.669 | 1.123 | 1.396 | 0.928 | 0.142 | **0.750** | **0.986** | **1.073** | **0.705** | **0.307** | **0.432** | **1.056** | **0.460** |

Note: 1. The statistical indicators of the best performing machine learning model of this climate zone in each season are highlighted in colour, and the statistical indicators of the better performing machine learning models are highlighted in grey.

2. The statistical indicators of the best performing machine learning models for each site in each season are highlighted in grey, and the statistical indicators of the better performing machine learning models are shown in bold.

**Table 19**

The mean statistics of the performance indicators for 1-7 days lead time $ET_o$ prediction using 5 methods at 4 stations in the SAR climate region during the four seasons of 2020-2021.

| Stations/Methods | Spring | | | | Summer | | | | Fall | | | | Winter | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | MAE (°C) | RMSE (°C) | RM | R | MAE (°C) | RMSE (°C) | RM | R | MAE (°C) | RMSE (°C) | RM | R | MAE (°C) | RMSE (°C) | RM | R |
| **SAR/Yanchi** | | | | | | | | | | | | | | | | |
| MLP$_p$ | 0.797 | 1.015 | 0.841 | 0.679 | **1.017** | **1.270** | **0.927** | **0.210** | 0.755 | 0.991 | 1.157 | 0.761 | 0.380 | 0.469 | 1.271 | 0.557 |
| XGBoost$_p$ | **0.805** | **1.042** | **0.828** | **0.680** | **1.114** | **1.368** | **0.878** | **0.203** | **0.714** | **0.959** | **0.937** | **0.699** | **0.303** | **0.420** | **1.087** | **0.518** |
| LightGBM$_p$ | **0.770** | **0.999** | **0.875** | **0.664** | 1.012 | 1.281 | 0.929 | 0.195 | 0.749 | 1.001 | 1.029 | 0.684 | 0.351 | 0.459 | 1.201 | 0.513 |
| CatBoost1$_p$ | 0.869 | 1.093 | 0.790 | 0.678 | 1.153 | 1.403 | 0.875 | 0.209 | **0.680** | **0.903** | **1.065** | **0.749** | **0.327** | **0.433** | **1.150** | **0.556** |
| CatBoost2 | 0.977 | 1.220 | 0.747 | 0.620 | 1.250 | 1.522 | 0.864 | 0.175 | 0.702 | 0.925 | 1.013 | 0.736 | 0.307 | 0.430 | 1.086 | 0.527 |
| **SAR/Haiyuan** | | | | | | | | | | | | | | | | |
| MLP$_p$ | **0.971** | **1.285** | **0.810** | **0.575** | **1.225** | **1.516** | **0.968** | **0.114** | 0.899 | 1.164 | 1.199 | 0.634 | 0.360 | 0.486 | 1.118 | 0.528 |
| XGBoost$_p$ | **0.980** | **1.286** | **0.811** | **0.577** | **1.225** | **1.509** | **0.964** | **0.121** | 0.881 | 1.132 | 1.058 | 0.562 | 0.362 | 0.492 | 1.114 | 0.518 |
| LightGBM$_p$ | 0.953 | 1.257 | 0.865 | 0.539 | 1.243 | 1.537 | 0.975 | 0.112 | 0.929 | 1.191 | 1.106 | 0.545 | **0.381** | **0.515** | **1.157** | **0.501** |
| CatBoost1$_p$ | 1.004 | 1.332 | 0.786 | 0.562 | 1.295 | 1.589 | 0.931 | 0.110 | **0.856** | **1.092** | **1.142** | **0.626** | **0.350** | **0.492** | **1.005** | **0.530** |
| CatBoost2 | 1.077 | 1.397 | 0.747 | 0.571 | 1.320 | 1.647 | 0.927 | 0.094 | **0.870** | **1.123** | **1.117** | **0.619** | 0.366 | 0.484 | 1.105 | 0.484 |
| **SAR/Tongxin** | | | | | | | | | | | | | | | | |
| MLP$_p$ | **0.916** | **1.201** | **0.854** | **0.639** | **1.333** | **1.639** | **0.975** | **0.163** | 1.038 | 1.355 | 1.194 | 0.629 | 0.400 | 0.494 | 1.197 | 0.611 |
| XGBoost$_p$ | 0.966 | 1.250 | 0.834 | 0.629 | **1.353** | **1.658** | **0.950** | **0.188** | **0.929** | **1.207** | **1.009** | **0.597** | 0.491 | 0.584 | 1.267 | 0.460 |
| LightGBM$_p$ | **0.907** | **1.181** | **0.915** | **0.629** | 1.389 | 1.770 | 1.046 | 0.177 | 1.084 | 1.443 | 1.129 | 0.562 | 0.401 | 0.507 | 1.151 | 0.537 |
| CatBoost1$_p$ | 0.955 | 1.252 | 0.831 | 0.630 | 1.410 | 1.739 | 0.967 | 0.160 | **0.983** | **1.276** | **1.136** | **0.620** | **0.358** | **0.492** | **1.066** | **0.595** |
| CatBoost2 | 1.097 | 1.401 | 0.746 | 0.630 | 1.492 | 1.828 | 0.902 | 0.171 | 0.938 | 1.240 | 1.046 | 0.611 | **0.358** | **0.481** | **1.052** | **0.562** |
| **SAR/Average** | | | | | | | | | | | | | | | | |
| MLP$_p$ | **0.895** | **1.167** | **0.835** | **0.631** | **1.192** | **1.475** | **0.957** | **0.162** | 0.897 | 1.170 | 1.183 | 0.675 | 0.380 | 0.483 | 1.195 | 0.565 |
| XGBoost$_p$ | 0.917 | 1.193 | 0.824 | 0.629 | **1.231** | **1.512** | **0.931** | **0.171** | **0.841** | **1.099** | **1.001** | **0.619** | 0.385 | 0.499 | 1.156 | 0.499 |
| LightGBM$_p$ | **0.877** | **1.146** | **0.885** | **0.611** | 1.215 | 1.529 | 0.983 | 0.161 | 0.921 | 1.212 | 1.088 | 0.597 | 0.378 | 0.494 | 1.170 | 0.517 |
| CatBoost1$_p$ | 0.943 | 1.226 | 0.802 | 0.623 | 1.286 | 1.577 | 0.924 | 0.160 | **0.840** | **1.090** | **1.114** | **0.665** | **0.345** | **0.472** | **1.074** | **0.560** |
| CatBoost2 | 1.050 | 1.339 | 0.747 | 0.607 | 1.354 | 1.666 | 0.898 | 0.147 | 0.837 | 1.096 | 1.059 | 0.665 | **0.344** | **0.465** | **1.081** | **0.524** |

Note: 1. The statistical indicators of the best performing machine learning model of this climate zone in each season are highlighted in colour, and the statistical indicators of the better performing machine learning models are highlighted in grey.

2. The statistical indicators of the best performing machine learning models for each site in each season are highlighted in grey, and the statistical indicators of the better performing machine learning models are shown in bold.

**Table 20**

The mean statistics of the performance indicators for 1-7 days lead time $ET_o$ prediction using 5 methods at 4 stations in the SHZ climate region during the four seasons of 2020-2021.

| Stations/Methods | Spring | Summer | Fall | Winter |
|---|---|---|---|---|

| | MAE (°C) | RMSE (°C) | RM | R | MAE (°C) | RMSE (°C) | RM | R | MAE (°C) | RMSE (°C) | RM | R | MAE (°C) | RMSE (°C) | RM | R |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **SHZ/Guyuan** | | | | | | | | | | | | | | | | |
| $MLP_p$ | **0.997** | **1.270** | **0.773** | **0.531** | **1.239** | **1.491** | **0.975** | **0.112** | 0.934 | 1.191 | 1.201 | 0.567 | **0.369** | **0.492** | **1.042** | **0.490** |
| $XGBoost_p$ | **1.022** | **1.278** | **0.758** | **0.567** | **1.245** | **1.480** | **0.918** | **0.134** | **0.927** | **1.142** | **1.200** | **0.584** | 0.374 | 0.508 | 1.028 | 0.461 |
| $LightGBM_p$ | 1.024 | 1.301 | 0.790 | 0.479 | 1.247 | 1.536 | 0.991 | 0.133 | 0.913 | 1.159 | 1.052 | 0.522 | **0.353** | **0.496** | **0.962** | **0.454** |
| $CatBoost1_p$ | 1.134 | 1.413 | 0.681 | 0.543 | 1.270 | 1.535 | 0.930 | 0.118 | **0.872** | **1.101** | **1.100** | **0.575** | 0.397 | 0.552 | 0.802 | 0.473 |
| CatBoost2 | 1.147 | 1.440 | 0.711 | 0.514 | 1.297 | 1.571 | 0.911 | 0.127 | 0.956 | 1.189 | 1.110 | 0.528 | 0.454 | 0.597 | 1.077 | 0.311 |
| **SHZ/Xiji** | | | | | | | | | | | | | | | | |
| $MLP_p$ | **0.574** | **0.764** | **0.871** | **0.648** | **0.913** | **1.116** | **1.007** | **0.057** | 0.713 | 0.935 | 1.204 | 0.690 | 0.203 | 0.252 | 1.199 | 0.638 |
| $XGBoost_p$ | **0.590** | **0.754** | **0.842** | **0.703** | **0.947** | **1.107** | **0.913** | **0.080** | **0.706** | **0.882** | **1.130** | **0.661** | 0.207 | 0.281 | 1.158 | 0.551 |
| $LightGBM_p$ | 0.580 | 0.767 | 0.886 | 0.638 | 0.964 | 1.168 | 0.962 | 0.038 | 0.677 | 0.901 | 1.064 | 0.633 | **0.150** | **0.201** | **1.039** | **0.613** |
| $CatBoost1_p$ | 0.601 | 0.783 | 0.829 | 0.674 | 0.927 | 1.133 | 0.968 | 0.069 | 0.692 | 0.883 | 1.162 | 0.676 | **0.175** | **0.243** | **1.043** | **0.621** |
| CatBoost2 | 0.693 | 0.890 | 0.768 | 0.632 | 0.958 | 1.154 | 0.911 | 0.093 | **0.654** | **0.852** | **1.062** | **0.666** | 0.181 | 0.242 | 0.882 | 0.583 |
| **SHZ/Average** | | | | | | | | | | | | | | | | |
| $MLP_p$ | **0.786** | **1.017** | **0.822** | **0.590** | **1.076** | **1.304** | **0.991** | **0.085** | 0.824 | 1.063 | 1.203 | 0.629 | **0.286** | **0.372** | **1.121** | **0.564** |
| $XGBoost_p$ | **0.806** | **1.016** | **0.800** | **0.635** | **1.096** | **1.294** | **0.916** | **0.107** | **0.817** | **1.012** | **1.165** | **0.623** | 0.291 | 0.395 | 1.093 | 0.506 |
| $LightGBM_p$ | 0.802 | 1.034 | 0.838 | 0.559 | 1.106 | 1.352 | 0.977 | 0.086 | 0.795 | 1.030 | 1.058 | 0.578 | **0.252** | **0.349** | **1.001** | **0.534** |
| $CatBoost1_p$ | 0.868 | 1.098 | 0.755 | 0.609 | 1.099 | 1.334 | 0.949 | 0.094 | **0.782** | **0.992** | **1.131** | **0.626** | 0.286 | 0.398 | 0.923 | 0.547 |
| CatBoost2 | 0.920 | 1.165 | 0.740 | 0.573 | 1.128 | 1.363 | 0.911 | 0.110 | 0.805 | 1.021 | 1.086 | 0.597 | 0.318 | 0.420 | 0.980 | 0.447 |

Note: 1. The statistical indicators of the best performing machine learning model of this climate zone in each season are highlighted in colour, and the statistical indicators of the better performing machine learning models are highlighted in grey.

2. The statistical indicators of the best performing machine learning models for each site in each season are highlighted in grey, and the statistical indicators of the better performing machine learning models are shown in bold.

## 3.5. Impact of weather variable forecasts from public weather forecasts on daily $ET_o$ forecasts

For all three climate zones, the performance of the four models developed based on public weather forecast data with a 1-day lead time was better than the performance of the four models developed based on daily observed meteorological data. To reliably and accurately analyse the weather forecast variables that cause daily $ET_o$ forecast errors, the models developed based on daily observed meteorological data were chosen to evaluate the impact of weather variables from public weather forecasts on the daily $ET_o$ forecasting performance. In this study, each of the four observed weather variables ($T_{max}$, $T_{min}$, SDun, and Wspd) was replaced in sequence with their corresponding forecast values with a 1-7 day lead time. This allowed for the identification of cases where a large change in the forecasted daily $ET_o$ indicates an error in the prediction resulting from the forecasted

https://doi.org/10.5194/hess-2023-158
Preprint. Discussion started: 31 July 2023
© Author(s) 2023. CC BY 4.0 License.


weather variable (Perera et al., 2014; Yang et al., 2016; Pelosi et al., 2016; Medina et al., 2018; Fan et al., 2021b). $T_{max}$, $T_{min}$, SDun, and Wspd were replaced by $T_{max}$, $T_{min}$, *SDun*, and *Wspd* in the public weather forecast 1-7 days ahead in turn to form the following new combinations: SC1 ($T_{max}$, $T_{min}$, SDun, and Wspd), SC2 ($T_{max}$, $T_{min}$, SDun, and Wspd), SC3 ($T_{max}$, $T_{min}$, *SDun*, and Wspd), and SC4 ($T_{max}$, $T_{min}$, SDun, and *Wspd*), all of which were composed of day-by-day observations for the combination denoted SC ($T_{max}$, $T_{min}$, SDun, and Wspd). The statistics of the mean values for the 1-7 day lead time $ET_o$ performance metrics using four models ($MLP_o$, $XGBoost_o$, $LightGBM_o$, and $CatBoost1_o$) for the three climate zones with the SC-SC4 input combination are shown in Tables 21-23.

First, for all models (except the LightGBM model at station ZW and the CatBoost1 model at station HN), the contribution of the public weather forecast variables to the error in the predicted daily $ET_o$ decreased in the order of *Wspd*, *SDun*, $T_{max}$, and $T_{min}$ for the AR (arid zone) climate zone, which is consistent with previous findings (Yang et al. 2016). Second, for all models (except the MLP model at station YAC, the CatBoost1 model at station TX, the MLP model at station XJ, the XGBoost model, and the LightGBM model), the contributions of the public weather forecast variables to the errors in the predicted daily $ET_o$ decreased in the order of *SDun*, *Wspd*, $T_{max}$, and $T_{min}$ and *SDun*, $T_{max}$, *Wspd*, and $T_{min}$ for the SAR (semiarid) and SHZ (semihumid zone) climate zones, which is consistent with the results of previous studies (Pelosi et al., 2016; Yang et al., 2016; Medina et al., 2018; Fan et al., 2021b).

These results indicate that for the study sites in the AR climate zone (arid zone), the main source of error in daily $ET_o$ prediction is Wspd transformed from the wind scale in public weather forecasts. For the study sites in the SAR (semiarid zone) and SHZ (semihumid zone) climate zones, SDun

https://doi.org/10.5194/hess-2023-158
Preprint. Discussion started: 31 July 2023
© Author(s) 2023. CC BY 4.0 License.


converted from the weather type of public weather forecasts contributes the most to the predicted daily $ET_o$ errors. First, the spatial variability of Wspd and DSun ($R_s$) due to the topography, elevation, distance, and cloudiness of the location of the study site (Yuan et al., 2015; Fick and Hijmasn, 2017; Li and Zha, 2018; Fan et al., 2021b) makes Wspd and DSun ($R_s$) the most difficult parameters to forecast accurately (Yang et al., 2016, Ballesteros et al., 2016); second, according to Cai et al. (2007), it is appropriate to estimate Wspd from wind scale predicted by public weather forecasts, but this estimation error is larger for arid regions with a high range of wind speed values. George et al. (1985) reported that the largest difference between predicted and measured reference crop evapotranspiration came from erroneous predictions of mean wind speed. Li and Beswick (2005) also reported that wind speed is a more serious source of error than solar radiation in estimating $ET_o$. In a study by Popova et al. (2005), it was noted that the effect of wind speed on $ET_o$ results was relatively small except in arid and windy areas.

As shown in Table 3, SDun was estimated using the sunshine hour coefficients derived from the 2004 measured solar radiation data from Daxing District, Beijing, using Equation (3). It was found that applying the sunshine hour coefficients derived from one region to other regions with different climate types will result in different degrees of error due to the climatic differences between regions. Perera et al. (2014) found that the largest source of error between predicted and observed $ET_o$ is the predictive performance of daily incoming solar radiation, followed by air temperature, dew point temperature, and wind speed for all advanced periods. Pelosi et al. (2016) indicated that the solar radiation forecast error has the greatest impact on the $ET_o$ forecast performance, followed by relative humidity and wind speed. The results of Medina et al. (2018) also indicated that the errors in solar radiation forecasts have the greatest impact on $ET_o$ forecasts, followed by errors in wind forecasts.

https://doi.org/10.5194/hess-2023-158
Preprint. Discussion started: 31 July 2023
© Author(s) 2023. CC BY 4.0 License.

Fan et al. (2021b) also showed that the contribution of predicted weather variables to the daily $ET_o$ error in all studied climate zones, such as the temperate continental zone (TCZ)/temperate monsoon zone (TMZ), was determined by $R_s$ (solar radiation), $W_s$ (wind speed), $T_{max}$, RH (relative humidity), Tmin/Rs (solar radiation), RH (relative humidity), $W_s$ (wind speed), $T_{max}$, and $T_{min}$ in decreasing order.

**Table 21**

Mean statistics of the $ET_o$ performance index for 1-7 days of lead time predicted by five models for the AR climate zone with five input combinations when replacing observed weather variables with weather variables predicted by public weather forecasts one by one.

| Stations/Methods | SC | | | | SC1 | | | | SC2 | | | | SC3 | | | | SC4 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | MAE (mm $d^{-1}$) | RMSE (mm $d^{-1}$) | RM | R | MAE (mm $d^{-1}$) | RMSE (mm $d^{-1}$) | RM | R | MAE (mm $d^{-1}$) | RMSE (mm $d^{-1}$) | RM | R | MAE (mm $d^{-1}$) | RMSE (mm $d^{-1}$) | RM | R | MAE (mm $d^{-1}$) | RMSE (mm $d^{-1}$) | RM | R |
| HN/MLP | 0.39 | 0.51 | 0.95 | 0.97 | 0.43 | 0.58 | 0.95 | 0.96 | 0.40 | 0.53 | 0.95 | 0.97 | **0.77** | **1.02** | **0.84** | **0.90** | **0.76** | **1.07** | **1.02** | **0.86** |
| YC/MLP | 0.31 | 0.41 | 0.95 | 0.98 | 0.37 | 0.51 | 0.95 | 0.96 | 0.29 | 0.40 | 0.96 | 0.98 | **0.59** | **0.80** | **0.89** | **0.92** | **0.68** | **1.04** | **1.14** | **0.87** |
| ZW/MLP | 0.41 | 0.55 | 0.97 | 0.96 | 0.48 | 0.64 | 0.97 | 0.94 | 0.41 | 0.55 | 0.98 | 0.96 | **0.65** | **0.88** | **0.91** | **0.90** | **0.74** | **1.07** | **1.03** | **0.84** |
| ZN/MLP | 0.35 | 0.49 | 0.93 | 0.97 | 0.43 | 0.60 | 0.94 | 0.95 | 0.34 | 0.48 | 0.95 | 0.97 | **0.62** | **0.85** | **0.90** | **0.90** | **0.66** | **0.96** | **1.06** | **0.88** |
| **AR/MLP** | 0.37 | 0.49 | 0.95 | 0.97 | 0.43 | 0.59 | 0.95 | 0.95 | 0.36 | 0.49 | 0.96 | 0.97 | **0.65** | **0.89** | **0.89** | **0.91** | **0.71** | **1.04** | **1.06** | **0.86** |
| HN/XGBoost | 0.35 | 0.47 | 0.98 | 0.97 | 0.42 | 0.57 | 0.98 | 0.96 | 0.36 | 0.48 | 0.98 | 0.97 | **0.65** | **0.88** | **0.90** | **0.92** | **0.66** | **0.89** | **1.03** | **0.90** |
| YC/XGBoost | 0.30 | 0.40 | 0.97 | 0.98 | 0.35 | 0.49 | 0.96 | 0.97 | 0.29 | 0.40 | 0.97 | 0.98 | **0.58** | **0.79** | **0.90** | **0.92** | **0.63** | **0.91** | **1.11** | **0.91** |
| ZW/XGBoost | 0.42 | 0.56 | 0.97 | 0.96 | 0.46 | 0.63 | 0.97 | 0.94 | 0.42 | 0.57 | 0.98 | 0.96 | **0.66** | **0.88** | **0.92** | **0.89** | **0.63** | **0.84** | **0.99** | **0.90** |
| ZN/XGBoost | 0.34 | 0.47 | 0.95 | 0.97 | 0.43 | 0.60 | 0.95 | 0.95 | 0.33 | 0.46 | 0.96 | 0.97 | **0.59** | **0.81** | **0.92** | **0.91** | **0.64** | **0.91** | **1.06** | **0.89** |
| **AR/XGBoost** | 0.35 | 0.47 | 0.97 | 0.97 | 0.42 | 0.57 | 0.97 | 0.95 | 0.35 | 0.48 | 0.97 | 0.97 | **0.62** | **0.84** | **0.91** | **0.91** | **0.64** | **0.89** | **1.05** | **0.90** |
| HN/LightGBM | 0.44 | 0.58 | 0.98 | 0.96 | 0.51 | 0.68 | 0.97 | 0.94 | 0.44 | 0.59 | 0.98 | 0.96 | **0.67** | **0.89** | **0.90** | **0.92** | **0.68** | **0.87** | **1.02** | **0.90** |
| YC/LightGBM | 0.32 | 0.42 | 0.96 | 0.98 | 0.37 | 0.52 | 0.96 | 0.96 | 0.33 | 0.43 | 0.97 | 0.97 | **0.59** | **0.79** | **0.89** | **0.92** | **0.64** | **0.91** | **1.11** | **0.89** |
| ZW/LightGBM | 0.50 | 0.65 | 0.98 | 0.94 | 0.58 | 0.78 | 0.98 | 0.91 | 0.50 | 0.66 | 0.98 | 0.94 | **0.67** | **0.88** | **0.93** | **0.91** | **0.64** | **0.84** | **0.99** | **0.90** |
| ZN/LightGBM | 0.35 | 0.47 | 0.95 | 0.97 | 0.44 | 0.60 | 0.95 | 0.55 | 0.34 | 0.46 | 0.96 | 0.97 | **0.59** | **0.81** | **0.92** | **0.91** | **0.66** | **0.96** | **1.06** | **0.88** |
| **AR/LightGBM** | 0.40 | 0.53 | 0.97 | 0.96 | 0.48 | 0.65 | 0.97 | 0.94 | 0.40 | 0.53 | 0.97 | 0.96 | **0.63** | **0.84** | **0.91** | **0.92** | **0.66** | **0.89** | **1.05** | **0.89** |
| HN/CatBoost1 | 0.44 | 0.57 | 0.97 | 0.96 | 0.50 | 0.67 | 0.97 | 0.95 | 0.45 | 0.60 | 0.97 | 0.96 | **0.70** | **0.93** | **0.89** | **0.92** | **0.66** | **0.86** | **1.01** | **0.91** |
| YC/CatBoost1 | 0.37 | 0.49 | 0.98 | 0.96 | 0.45 | 0.61 | 0.98 | 0.94 | 0.38 | 0.50 | 0.98 | 0.96 | **0.57** | **0.76** | **0.91** | **0.93** | **0.62** | **0.81** | **1.10** | **0.91** |
| ZW/CatBoost1 | 0.47 | 0.62 | 0.97 | 0.95 | 0.54 | 0.74 | 0.97 | 0.92 | 0.47 | 0.62 | 0.97 | 0.95 | **0.66** | **0.86** | **0.92** | **0.91** | **0.65** | **0.87** | **0.99** | **0.89** |
| ZN/CatBoost1 | 0.39 | 0.52 | 0.97 | 0.96 | 0.51 | 0.72 | 0.97 | 0.92 | 0.39 | 0.53 | 0.97 | 0.96 | **0.58** | **0.78** | **0.93** | **0.92** | **0.61** | **0.82** | **1.05** | **0.90** |
| **AR/CatBoost1** | 0.42 | 0.56 | 0.97 | 0.96 | 0.50 | 0.68 | 0.97 | 0.93 | 0.42 | 0.56 | 0.97 | 0.96 | **0.63** | **0.83** | **0.91** | **0.92** | **0.64** | **0.84** | **1.04** | **0.90** |

Note: The statistical indicators with the largest error contribution to the predicted daily $ET_o$ of each site (climate zone) are highlighted in blue, and the statistical indicators with the second largest error contribution to the predicted daily $ET_o$ of each site (climate zone) are highlighted in grey.

**Table 22**

Mean statistics of the $ET_o$ performance index for 1-7 days of lead time predicted by five models for the SAR climate zone with five input combinations when replacing observed weather variables with weather variables predicted by public weather forecasts one by one.

| Stations/Methods | SC | | | | SC1 | | | | SC2 | | | | SC3 | | | | SC4 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | MAE | RMSE | RM | R | MAE | RMSE | RM | R | MAE | RMSE | RM | R | MAE | RMSE | RM | R | MAE | RMSE | RM | R |

https://doi.org/10.5194/hess-2023-158
Preprint. Discussion started: 31 July 2023
© Author(s) 2023. CC BY 4.0 License.



| | (mm d$^{-1}$) | (mm d$^{-1}$) | | | (mm d$^{-1}$) | (mm d$^{-1}$) | | | (mm d$^{-1}$) | (mm d$^{-1}$) | | | (mm d$^{-1}$) | (mm d$^{-1}$) | | | (mm d$^{-1}$) | (mm d$^{-1}$) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| YAC/MLP | 0.45 | 0.56 | 1.03 | 0.96 | 0.50 | 0.63 | 1.02 | 0.95 | 0.50 | 0.62 | 1.04 | 0.95 | 0.75 | 1.00 | 0.90 | 0.89 | 0.45 | 0.56 | 1.03 | 0.96 |
| HY/MLP | 0.36 | 0.47 | 1.00 | 0.97 | 0.47 | 0.63 | 0.99 | 0.94 | 0.36 | 0.48 | 1.01 | 0.97 | 0.65 | 0.90 | 0.94 | 0.88 | 0.63 | 0.91 | 1.03 | 0.88 |
| TX/MLP | 0.45 | 0.63 | 0.96 | 0.96 | 0.61 | 0.84 | 0.95 | 0.93 | 0.45 | 0.62 | 0.97 | 0.96 | 0.68 | 0.94 | 0.93 | 0.91 | 0.66 | 0.92 | 0.96 | 0.91 |
| **SAR/MLP** | 0.42 | 0.55 | 1.00 | 0.96 | 0.53 | 0.70 | 0.99 | 0.94 | 0.44 | 0.57 | 1.00 | 0.96 | 0.69 | 0.95 | 0.92 | 0.89 | 0.58 | 0.79 | 1.01 | 0.92 |
| YAC/XGBoost | 0.32 | 0.41 | 1.00 | 0.98 | 0.44 | 0.60 | 0.98 | 0.96 | 0.33 | 0.42 | 1.00 | 0.98 | 0.67 | 0.94 | 0.88 | 0.92 | 0.52 | 0.66 | 1.04 | 0.94 |
| HY/XGBoost | 0.37 | 0.50 | 0.98 | 0.96 | 0.46 | 0.63 | 0.98 | 0.94 | 0.37 | 0.50 | 0.99 | 0.96 | 0.69 | 0.94 | 0.93 | 0.87 | 0.55 | 0.75 | 1.01 | 0.91 |
| TX/XGBoost | 0.45 | 0.63 | 0.97 | 0.96 | 0.60 | 0.83 | 0.96 | 0.93 | 0.46 | 0.63 | 0.97 | 0.96 | 0.68 | 0.93 | 0.94 | 0.91 | 0.65 | 0.90 | 0.97 | 0.91 |
| **SAR/XGBoost** | 0.38 | 0.51 | 0.99 | 0.97 | 0.50 | 0.69 | 0.97 | 0.94 | 0.38 | 0.52 | 0.99 | 0.97 | 0.68 | 0.93 | 0.92 | 0.90 | 0.57 | 0.77 | 1.01 | 0.92 |
| YAC/LightGBM | 0.37 | 0.48 | 0.99 | 0.97 | 0.50 | 0.67 | 0.97 | 0.94 | 0.37 | 0.49 | 0.99 | 0.97 | 0.66 | 0.89 | 0.88 | 0.93 | 0.54 | 0.68 | 1.03 | 0.94 |
| HY/LightGBM | 0.35 | 0.48 | 0.99 | 0.97 | 0.47 | 0.65 | 0.97 | 0.94 | 0.36 | 0.49 | 0.99 | 0.97 | 0.67 | 0.93 | 0.93 | 0.87 | 0.54 | 0.75 | 1.00 | 0.91 |
| TX/LightGBM | 0.46 | 0.64 | 0.97 | 0.96 | 0.62 | 0.85 | 0.96 | 0.92 | 0.46 | 0.64 | 0.97 | 0.96 | 0.68 | 0.93 | 0.93 | 0.92 | 0.66 | 0.90 | 0.98 | 0.91 |
| **SAR/LightGBM** | 0.40 | 0.53 | 0.98 | 0.97 | 0.53 | 0.72 | 0.97 | 0.93 | 0.40 | 0.54 | 0.98 | 0.96 | 0.67 | 0.91 | 0.91 | 0.91 | 0.58 | 0.78 | 1.00 | 0.92 |
| YAC/CatBoost1 | 0.35 | 0.45 | 0.99 | 0.97 | 0.43 | 0.57 | 0.98 | 0.96 | 0.37 | 0.47 | 0.99 | 0.97 | 0.71 | 0.95 | 0.86 | 0.91 | 0.54 | 0.67 | 1.04 | 0.94 |
| HY/CatBoost1 | 0.35 | 0.47 | 0.98 | 0.97 | 0.49 | 0.68 | 0.97 | 0.93 | 0.36 | 0.47 | 0.98 | 0.97 | 0.63 | 0.87 | 0.93 | 0.89 | 0.58 | 0.80 | 1.00 | 0.90 |
| TX/CatBoost1 | 0.51 | 0.68 | 0.97 | 0.95 | 0.68 | 0.94 | 0.96 | 0.90 | 0.51 | 0.68 | 0.97 | 0.95 | 0.70 | 0.94 | 0.92 | 0.92 | 0.68 | 0.90 | 0.97 | 0.91 |
| **SAR/CatBoost1** | 0.41 | 0.53 | 0.98 | 0.97 | 0.54 | 0.73 | 0.97 | 0.93 | 0.41 | 0.54 | 0.98 | 0.96 | 0.68 | 0.92 | 0.90 | 0.91 | 0.60 | 0.79 | 1.00 | 0.92 |

Note: The statistical indicators with the largest error contribution to the predicted daily $ET_o$ of each site (climate zone) are highlighted in blue, and the statistical indicators with the second largest error contribution to the predicted daily $ET_o$ of each site (climate zone) are highlighted in grey.

**Table 23**

Mean statistics of the $ET_o$ performance index for 1-7 days of lead time predicted by five models for the SHZ climate zone with five input combinations when replacing observed weather variables with weather variables predicted by public weather forecasts one by one.

| Stations/Methods | SC | | | | SC1 | | | | SC2 | | | | SC3 | | | | SC4 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | MAE (mm d$^{-1}$) | RMSE (mm d$^{-1}$) | RM | R | MAE (mm d$^{-1}$) | RMSE (mm d$^{-1}$) | RM | R | MAE (mm d$^{-1}$) | RMSE (mm d$^{-1}$) | RM | R | MAE (mm d$^{-1}$) | RMSE (mm d$^{-1}$) | RM | R | MAE (mm d$^{-1}$) | RMSE (mm d$^{-1}$) | RM | R |
| GY/MLP | 0.39 | 0.51 | 1.00 | 0.95 | 0.58 | 0.78 | 1.01 | 0.89 | 0.39 | 0.51 | 1.00 | 0.95 | 0.67 | 0.90 | 0.96 | 0.85 | 0.54 | 0.77 | 0.99 | 0.89 |
| XJ/MLP | 0.25 | 0.33 | 0.98 | 0.97 | 0.32 | 0.44 | 0.97 | 0.95 | 0.26 | 0.35 | 1.00 | 0.97 | 0.51 | 0.71 | 0.97 | 0.87 | 0.57 | 0.92 | 1.19 | 0.85 |
| **SHZ/MLP** | 0.32 | 0.42 | 0.99 | 0.96 | 0.45 | 0.61 | 0.99 | 0.92 | 0.32 | 0.43 | 1.00 | 0.96 | 0.59 | 0.80 | 0.96 | 0.86 | 0.55 | 0.85 | 1.09 | 0.87 |
| GY/XGBoost | 0.39 | 0.52 | 1.03 | 0.95 | 0.57 | 0.77 | 1.03 | 0.89 | 0.40 | 0.53 | 1.03 | 0.95 | 0.67 | 0.89 | 1.00 | 0.85 | 0.52 | 0.72 | 1.00 | 0.91 |
| XJ/XGBoost | 0.27 | 0.34 | 1.00 | 0.97 | 0.33 | 0.45 | 0.97 | 0.95 | 0.27 | 0.36 | 1.01 | 0.97 | 0.50 | 0.68 | 0.98 | 0.88 | 0.37 | 0.47 | 1.10 | 0.96 |
| **SHZ/XGBoost** | 0.33 | 0.43 | 1.02 | 0.96 | 0.45 | 0.61 | 1.00 | 0.92 | 0.34 | 0.44 | 1.02 | 0.96 | 0.59 | 0.79 | 0.99 | 0.86 | 0.45 | 0.60 | 1.05 | 0.93 |
| GY/LightGBM | 0.41 | 0.54 | 1.03 | 0.95 | 0.57 | 0.76 | 1.03 | 0.89 | 0.42 | 0.54 | 1.03 | 0.95 | 0.69 | 0.91 | 0.99 | 0.84 | 0.52 | 0.71 | 0.99 | 0.91 |
| XJ/LightGBM | 0.27 | 0.35 | 1.00 | 0.97 | 0.34 | 0.47 | 0.98 | 0.95 | 0.27 | 0.36 | 1.01 | 0.97 | 0.49 | 0.68 | 0.97 | 0.88 | 0.38 | 0.49 | 1.11 | 0.96 |
| **SHZ/LightGBM** | 0.34 | 0.44 | 1.01 | 0.96 | 0.46 | 0.62 | 1.00 | 0.92 | 0.34 | 0.45 | 1.02 | 0.96 | 0.59 | 0.79 | 0.98 | 0.86 | 0.45 | 0.60 | 1.05 | 0.93 |
| GY/CatBoost1 | 0.47 | 0.60 | 1.01 | 0.94 | 0.64 | 0.86 | 1.02 | 0.86 | 0.48 | 0.61 | 1.01 | 0.94 | 0.69 | 0.89 | 0.96 | 0.86 | 0.51 | 0.68 | 0.96 | 0.92 |
| XJ/CatBoost1 | 0.30 | 0.38 | 1.01 | 0.96 | 0.42 | 0.57 | 0.99 | 0.92 | 0.31 | 0.39 | 1.00 | 0.96 | 0.45 | 0.60 | 0.97 | 0.91 | 0.40 | 0.48 | 1.10 | 0.95 |
| **SHZ/CatBoost1** | 0.39 | 0.49 | 1.01 | 0.95 | 0.53 | 0.72 | 1.00 | 0.89 | 0.39 | 0.50 | 1.01 | 0.95 | 0.57 | 0.74 | 0.97 | 0.89 | 0.45 | 0.58 | 1.03 | 0.94 |

Note: The statistical indicators with the largest error contribution to the predicted daily $ET_o$ of each site (climate zone) are highlighted in blue, and the statistical indicators with the second largest error contribution to the predicted daily $ET_o$ of each site (climate zone) are highlighted in grey.

## 4. Conclusions

https://doi.org/10.5194/hess-2023-158
Preprint. Discussion started: 31 July 2023
© Author(s) 2023. CC BY 4.0 License.
Hydrology and Earth System Sciences
Discussions

In this study, public weather forecasts with a 1-7 day lead time were used to compare the performance of five models developed based on public weather forecast data with a 1-day lead time and four models developed based on daily observed meteorological data in predicting daily ET$_o$ at nine stations in three climatic regions of Ningxia, China. The forecast performance of weather variables in public weather forecasting (2014-2021) and daily ET$_o$ predicted by these nine models were analysed and evaluated on a daily scale, and the forecast performance of the weather variables in the public weather forecast (2020-2021) and daily ET$_o$ predicted by the five models developed based on the public weather forecast with a 1-day lead time were analysed and evaluated in terms of seasonality. The optimal input combination for each machine learning model was determined, and the weather forecast variables contributing to the error in the model predicted daily ET$_o$ were identified. The main conclusions of this study are as follows:

First, for the three climate zones, the performance of the four models developed based on public weather forecast data with a 1-day lead time was better than that of the four models developed based on daily observation meteorological data with corresponding input combinations. When category data such as wind scale (WS) and weather type (WT1 and WT2) were added directly to the input combinations of the CatBoost2 model, the performance of this model in predicting daily ET$_o$ was lower than that of the CatBoost1p model in the testing period; that is, the performance of the CatBoost2 model in terms of predicting daily ET$_o$ did not improve during the testing period.

Second, the performance of the five models, MLP$_p$, XGBoost$_p$, LightGBM$_p$, CatBoost1$_p$ and CatBoost2, in terms of daily ET$_o$ prediction, was highest in winter and the lowest in summer for all three climate zones. In terms of predicting daily ET$_o$ with a 1-7 day lead time in all seasons, XGBoost$_p$ with C2 as input, MLP$_p$ with C4 as input, CatBoost2 with C8 as input and LightGBM$_p$

https://doi.org/10.5194/hess-2023-158
Preprint. Discussion started: 31 July 2023
© Author(s) 2023. CC BY 4.0 License.
Hydrology and Earth System Sciences
Discussions
Open Access

with C1 as input are recommended as the best models for spring, summer, fall and winter in the AR climate zone, respectively. $LightGBM_p$ with C2 as input, $MLP_p$ with C4 as input, $CatBoost1_p$ with C3 as input and CatBoost2 with C8 as input are recommended as the best models for spring, summer, fall and winter in the SAR climate zone, respectively. $XGBoost_p$ with C4 as input, $CatBoost1_p$ with C3 as input and $LightGBM_p$ with C1 as input are recommended as the best models for spring and summer, fall and winter in the SHZ climate zone, respectively.

Finally, for the AR climate zone (arid zone), the contribution of the weather variables in the public weather forecast to the error in the predicted daily $ET_o$ decreased in the order of *Wspd*, *SDun*, $T_{max}$, and $T_{min}$; for the SAR climate zone (semiarid zone), the contribution of the weather variables in the public weather forecast to the error in the predicted daily $ET_o$ decreased in the order of SDun, *Wspd*, $T_{max}$, and $T_{min}$; and for the SHZ climate zone (semihumid zone), the contribution of the weather variables in the public weather forecast to the predicted daily $ET_o$ decreased in the order of *SDun*, $T_{max}$, *Wspd*, and $T_{min}$.

In addition, in terms of the daily scale performance of weather variables in public weather forecasts, the forecast performance follows a decreasing order of $T_{min}$>$T_{max}$>SDun>Wspd. In the seasonal analysis (2020-2021) of the weather variable forecast performance for three climate zones, the average performance for $T_{max}$ is in the order of summer (fall) > fall (summer) > winter > spring. For $T_{min}$, the order is fall (summer) > summer (fall) > winter (spring) > spring (winter), and for SDun, the order is winter > spring > fall > summer. Thus, the average performance of the forecasts decreases sequentially.

© Author(s) 2023. CC BY 4.0 License.


**Declaration of Competing Interest**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Acknowledgements**

This study was supported by the Ningxia Key Research and Development Program (Special Talents) (2019BEB04040), the Natural Science Foundation of Ningxia (Project No. 2020AAC03039), and the First class Discipline Construction Project of Ningxia (No. NXYLXK2021A03). The observed meteorological data obtained from the China Meteorological Data Sharing Service (http://data.cma.gov.cn) and the weather forecast data obtained from the China Weather Network (http://www.weather.com.cn) are gratefully acknowledged. The hard work of the reviewers for this paper is also greatly appreciated.

**References**

Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998. Crop evapotranspiration: guidelines for computing crop water requirements. FAO Irrigation and Drainage Paper No. 56. FAO, Rome, Italy.

Antonopoulos V Z, Antonopoulos A V, 2017. Daily reference evapotranspiration estimates by artificial neural networks technique and empirical equations using limited input climate variables. Comput. Electron. Agric. 132: 86-96.

Ballesteros, R., Ortega, J.F., Moreno, M.Á., 2016. FORETo: new software for reference evapotranspiration forecasting. J. Arid Environ. 124, 128–141.--- FORETo:

Cai, J.B., Liu, Y., Lei, T., Pereira, L.S., 2007. Estimating reference evapotranspiration with the FAO Penman–Monteith equation using daily weather forecast messages. Agric. For. Meteorol. 145 (1), 22–35.

Cai J B, Liu Y, Xu D, et al, 2009. Simulation of the soil water balance of wheat using daily weather forecast messages to estimate the reference evapotranspiration. Hydrol. Earth Sys Sci . 13(7): 1045-1059.

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 pp 785-794.

Despotovic, M., Nedic, V., Despotovic, D., Cvetanovic, S., 2015. Review and statistical analysis of different global solar radiation sunshine models. Renew. Sustain. Energy Rev. 52, 1869–1880.

Dorogush AV, Ershov V, Gulin A (2018) CatBoost: gradient boosting with categorical features support

Elbeltagi A, Nagy A, Mohammed S, et al, 2022. Combination of limited meteorological data for predicting reference crop evapotranspiration using artificial neural network method. Agronomy. 12(2): 516.

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

Fan J, Ma X, Wu L, et al, 2019. Light Gradient Boosting Machine: An efficient soft computing model for estimating daily reference evapotranspiration with local and external meteorological data. Agric.Water Manage. 225: 105758.

Fan, J., Zheng, J., Wu, L., Zhang, F., 2021a. Estimation of daily maize transpiration using support vector machines, extreme gradient boosting, artificial and deep neural networks models. Agric.Water Manage. 245, 106547.

Fan J, Wu L, Zheng J, et al, 2021b. Medium-range forecasting of daily reference evapotranspiration across China using numerical weather prediction outputs downscaled by extreme gradient boosting. J. Hydrol. 601: 126664.

Feng, Y., Cui, N., Gong, D., Zhang, Q., Zhao, L., 2017. Evaluation of random forests and generalized regression neural networks for daily reference evapotranspiration modelling. Agric. Water Manage. 193, 163–173.

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

Ferreira, L.B., da Cunha, F.F., de Oliveira, R.A., Fernandes Filho, E.I., 2019. Estimation of reference evapotranspiration in Brazil with limited meteorological data using ANN and SVM – a new approach. J. Hydrol. 572, 556–570.

Fick, S.E., Hijmans, R.J., 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. Int. J. Climatol. 37 (12), 4302–4315.

GB/T 35227—2017, Specifications for surface meteorological observation—Wind direction and wind speed[S].Beijing: China Standard Press. 2017.

George, W., Pruitt, W.O., Dong, A., 1985. Evapotranspiration modelling In: CIMIS final report 10013-A. Land Air Water Res. Paper Ser. B 53812, 3.36–3.61.

Hinton, G. E., Osindero, S., & Teh, Y. W., 2006. A fast learning algorithm for deep belief nets. Neural computation, 2006, 18(7): 1527-1554.

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

Irmak, S., Irmak, A., Allen, R. G., Jones, J. W., 2003. Solar and Net Radiation-Based Equations to Estimate Reference Evapotranspiration in Humid Climates. J. Irrig. Drain Eng. 129(5), 336–347.

Jiang, S., Liang, C., Cui, N., Zhao, L., Du, T., Hu, X., Feng, Y., 2019. Impacts of climatic variables on reference evapotranspiration during growing season in Southwest China. Agric. Water Manage. 216, 365–378.

Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q, Liu T (2017) LightGBM: a highly efficient gradient boosting decision tree. 31st Conference on Neural Information Processing Systems

Kisi O, Zounemat-Kermani M (2014) Comparison of two different adaptive neuro-fuzzy inference systems in modelling daily reference evapotranspiration. Water Resour Manag 28(9):2655–2675

Kottek M, Grieser J, Beck C, Rudolf B, Rubel F (2006) World map of the Köppen-Geiger climate classification updated. Meteorol Z 15(3):259–263

Landeras G, Ortiz-Barredo A, López J J, 2008. Comparison of artificial neural network models and empirical and semi-empirical equations for daily reference evapotranspiration estimation in the Basque Country (Northern Spain). Agric. Water Manage. 95(5): 553-565.

Li D, Chen J, Luo Y, et al, 2018. Short-term daily forecasting of crop evapotranspiration of rice using public weather forecasts. Paddy and Water Environment. 16: 397-410.

Li, F.Z., Beswick, A., 2005. Sensitivity of the FAO-56 crop reference evapotranspiration to different input data. Technical report. Queensland Government, Natural Resources and Mines, pp. 1–14.

Li, L., Zha, Y., 2018. Mapping relative humidity, average and extreme temperature in hot summer over China. Sci. Total Environ. 615, 875–881.

Liu B, Liu M, Cui Y, et al, 2020. Assessing forecasting performance of daily reference evapotranspiration using public weather forecast and numerical weather prediction. J. Hydrol. 590: 125547.

Liu, Q., Wu, Z., Cui, N., Zhang, W., Wang, Y., et al, 2022. Genetic Algorithm-Optimized Extreme Learning Machine Model for Estimating

© Author(s) 2023. CC BY 4.0 License.


Daily Reference Evapotranspiration in Southwest China. Atmosphere. 13(6): 971.

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. Agric. Water Manage. 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. Resour Manage. 29, 3863–3876.

Mallikarjuna, P., Jyothy, S. A., Murthy, D. S., Reddy, K. C., 2014. Performance of recalibrated equations for the estimation of daily reference evapotranspiration. Water Resour Manag 28(13):4513–4535

Martınez-Cob A, Tejero-Juste M, 2015. A wind-based qualitative calibration of the Hargreaves ETo estimation equation in semiarid regions. Agric. water manage. 64(3): 251-264.

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

Medina H, Tian D, Srivastava P, et al, 2018. Medium-range reference evapotranspiration forecasts for the contiguous United States based on multi-model numerical weather predictions. J. Hydrol. 562: 502-517.

Medina H, Tian D, 2020. Comparison of probabilistic post-processing approaches for improving numerical weather prediction-based daily and weekly reference evapotranspiration forecasts. Hydrol. Earth Sys Sci. 24(2): 1011-1030.

Pelosi A, Medina H, Villani P, et al, 2016. Probabilistic forecasting of reference evapotranspiration with a limited area ensemble prediction system. Agric. Water Manage. 178: 106-118.

Perera, K.C., Andrew, W.W., Bandara, N., Biju, G., 2014. Forecasting daily reference evapotranspiration for Australia using numerical weather prediction outputs. Agric. Forest Meteorol. 194, 50–63.

Popova, Z., Kercheva, M. & Pereira, L. S. Validation of the FAO methodology for computing ETo with limited data, application to south Bulgaria. Irrig. Drain. 55 (2), 201–215.

Prokhorenkova L, Gusev G, Vorobev A et al. (2017) CatBoost: unbiased boosting with categorical features

Silva, D., Meza, F. J., Varas, E., 2010. Estimating reference evapotranspiration (ETo) using numerical weather forecast data in central Chile. J. Hydrol. 382(1-4), 64-71.

Tomas-Burguera, M., Vicente-Serrano, S.M., Grimalt, M., Begueria, S., 2017. Accuracy of reference evapotranspiration ($ET_o$) estimates under data scarcity scenarios in the Iberian peninsula. Agric. Water Manage. 182, 103–116.

Traore, S., Guven, A., 2013. New algebraic formulations of evapotranspiration extracted from gene-expression programming in the tropical seasonally dry regions of West Africa. Irrig. Sci. 31 (1), 1–10.

Traore S, Luo Y, Fipps G, 2016. Deployment of artificial neural network for short-term forecasting of evapotranspiration using public weather forecast restricted messages. Agric. Water Manage. 163: 363-379.

Traore S, Luo Y, Fipps G, 2017. Gene-expression programming for short-term forecasting of daily reference evapotranspiration using public weather forecast information. Water. Resour Manage. 31 (15), 4891-4908.

Vanella D, Intrigliolo D S, Consoli S, et al, 2020. Comparing the use of past and forecast weather data for estimating reference evapotranspiration. Agric. Forest Meteorol. 295: 108196.

Wu L, Peng Y, Fan J, et al, 2019a. Machine learning models for the estimation of monthly mean daily reference evapotranspiration based on cross-station and synthetic data. Hydrol. Res. 50(6): 1730-1750.

Wu L, Fan J, 2019b. Comparison of neuron-based, kernel-based, tree-based and curve-based machine learning models for predicting daily reference evapotranspiration. PloS one. 14(5): e0217520.

Yang, Y., Cui, Y., Luo, Y., Lyu, X., Traore, S., Khan, S., Wang, W., 2016. Short-term forecasting of daily reference evapotranspiration using the Penman-Monteith model and public weather forecasts. Agric. Water Manage. 177, 329–339.

Yang, Y., Cui, Y., Bai, K., Luo, T., Dai, J., Wang, W., Luo, Y., 2019a. Short-term forecasting of daily reference evapotranspiration using the reduced-set Penman-Monteith model and public weather forecasts. Agric. Water Manage. 211, 70–80.

Yang,Y., Luo,Y., Wu,C., Zheng,H., Zhang,L., Cui,Y., Sun,N., Wang,L., 2019b. Evaluation of six equations for daily reference evapotranspiration estimating using public weather forecast message for different climate regions across China. Agric.Water Manage. 222,

© Author(s) 2023. CC BY 4.0 License.


386–399.

Yin J, Deng Z, Ines A V M, et al, 2020. Forecast of short-term daily reference evapotranspiration under limited meteorological variables using a hybrid bi-directional long short-term memory model (Bi-LSTM). Agric. Water Manage. 242: 106386.

Yuan, W., Xu, B., Chen, Z., Xia, J., Xu, W., Chen, Y., Fu, Y., 2015. Validation of China-wide interpolated daily climate variables from 1960 to 2011. Theoret. Appl. Climatol. 119 (3), 689–700.

Zhang Y, Zhao Z, Zheng J, 2020. CatBoost: A new approach for estimating daily reference crop evapotranspiration in arid and semi-arid regions of Northern China. J. Hydrol. 588: 125087.

Zhou Z, Zhao L, Lin A, et al, 2020. Exploring the potential of deep factorization machine and various gradient boosting models in modeling daily reference evapotranspiration in China. Arab J. Geosci. 13: 1-20.