Research papers

Merging multiple satellite-based precipitation products and gauge observations using a novel double machine learning approach

Ling Zhang [a,*], Xin Li [b,c], Donghai Zheng [b], Kun Zhang [b], Qimin Ma [d], Yanbo Zhao [a], Yingchun Ge [a]

[a] *Key Laboratory of Remote Sensing of Gansu Province, Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences, Lanzhou 730000, China*
[b] *National Tibetan Plateau Data Center, Key Laboratory of Tibetan Environmental Changes and Land Surface Processes, Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing 100101, China*
[c] *CAS Center for Excellence in Tibetan Plateau Earth Sciences, Beijing 100101, China*
[d] *College of Resources and Environment, Chengdu University of Information Technology, Chengdu 610225, China*

ARTICLE INFO

ABSTRACT

This study proposed a novel double machine learning (DML) approach to merge multiple satellite-based precipitation products (SPPs) and gauge observations, and tested its reliability and validity over the Chinese mainland. The DML approach was mainly developed based on the classification model of random forest (RF) in combination with the regression models of the machine learning (ML) algorithms including RF, artificial neural network (ANN), support vector machine (SVM) and extreme learning machine (ELM). This led to four DML algorithms, i.e., RF-RF, RF-ANN, RF-SVM, and RF-LM. The performance of the DML algorithms were compared to the single machine learning (SML) algorithms developed based solely on the regression models of RF, ANN, SVM, and ELM, and to the liner merging methods including the inverse error variance weighting, the one-outlier-removed average, and the optimized weight average. In total, we produced twelve precipitation products including four of the DML algorithms, four of the SML algorithms, three of the liner merging methods, and another one generated via the gauge-only interpolation. The precipitation observations at 697 gauges were spatially and randomly divided into two parts (i.e., 70% and 30%), one was used for the training of the ML algorithms or for the interpolation, while the other for the performance evaluations. Results indicate that the DML algorithms outperform the other merging methods, the gauge-only interpolation, and the original SPPs over the Chinese mainland. The median Kling-Gupta efficiency (KGE) ranges 0.67–0.71 for the merged products of DML, which are obviously higher than the original SPPs (0.31–0.54), the linear merged product (0.54–0.55), gauge-only interpolated product (0.62), and the SML-based products (0.47–0.65). The DML-based products also exhibit better performances than the other products in detecting precipitation events with the threshold of 1 mm/day, and outperform the original SPPs regardless of the precipitation thresholds. Further analyses imply that: (i) the DML-based products could outperform the original SPPs even with a small training dataset size; (ii) the superiority of the DML approach to SML is mainly due to that the former can better capture the temporal dynamics of precipitation; (iii) the added values of the merged products of DML relative to the original SPPs and the gauge-only product vary with the sizes of the training dataset; and (iv) the ensemble of the DML algorithms could not further improve the accuracy of the precipitation estimates. This study not only provided an effective and robust tool for the fusion of multiple SPPs and gauge observations, but also, for the first time, compared the performance of various ML algorithms in merging satellite and gauge-based precipitation.

1. Introduction

Precipitation is one of the key components in the hydrological and atmospheric cycles. It is essential not only to the understanding of water balance and climate variability (Markonis et al., 2019; Du et al., 2020), but also to a wide range of applications such as flood forecasting (Breugem et al., 2020), drought monitoring (Zhong et al., 2019), water resource managements (Brodeur and Steinschneider, 2020), and

---

hydrological and earth system modeling (Tesfa et al., 2020; Zhang et al., 2020b). However, it is very challenging to obtain high-quality precipitation estimates at fine spatial–temporal resolutions (Xie and Xiong, 2011). Rain gauges are the traditional and reliable tools used to measure point-scale precipitation, which typically serve as the benchmark for the evaluation of various precipitation products. However, they are scarcely and unevenly distributed in most regions of the world (Sharifi et al., 2019; Beck et al., 2020); and the spatial continuous precipitation estimates based solely on gauge-based observations are subjected to large uncertainties (Baez-Villanueva-test et al., 2020).

Over the last few decades, the rapid development of remote sensing techniques provides an unprecedented opportunity for estimating spatial continuous precipitation at the global scale. Satellite-based precipitation products (SPPs) are increasingly available to the public, which significantly improved our understanding of the precipitation characteristics (Chen and Gao, 2018; Gupta et al., 2019; Markonis et al., 2019), and promoted a variety of hydrometeorological applications (Munier et al., 2014; Belabid et al., 2019; Pellet et al., 2019), especially over the gauge-sparse regions. Nevertheless, the SPPs are inherently subjected to large uncertainties and biases arising from the retrieval algorithms, the indirect measurements, and the deficiencies of the sensors (e.g., the false precipitation detections by infrared sensors if cold nonprecipitating clouds exist, and the undervalue of localized storm events by microwave sensors) (Bharti and Singh, 2015; Ebrahimi et al., 2017). They typically have trouble in estimating heavy and solid rainfall (Xu et al., 2017) as well as the local orographic precipitation (Bharti and Singh, 2015).

The gauge-based precipitation observations have high accuracy, but with limited coverages and uneven distributions. On the other hand, the SPPs have continuous and large coverages but with large biases. Hence, in recent years, many efforts have been made to merge SPPs and gauge observations to improve the accuracy and spatial coverage of the precipitation estimates (Nie et al., 2015; Yang et al., 2017; Chen et al., 2018; Beck et al., 2019; Baez-Villanueva et al., 2020; Xu et al., 2020). A variety of methods have been proposed to merge satellite and gauge-based precipitation, including (i) the simplest linear merging methods, e.g., the one-outlier-removed average (Shen et al., 2014), the inverse error variance weighting (Mastrantonas et al., 2019) and the inverse-root-mean-square-error weighting (Yang et al., 2017), (ii) the bias correction or residual-based methods, e.g., the probability-mapping method (Zhang and Tang, 2015), the geographical difference/ratio analysis (Duan and Bastiaanssen, 2013; Bai et al., 2019) and the kernel smoothing (Li and Shao, 2010), (iii) the intermediate-complexity approaches, e.g., the geographically weighted regression (GWR) (Chao et al., 2018) and the optimal interpolation technique (Xie and Xiong, 2011; Nie et al., 2015), and (iv) some other more complex methods such as the Bayesian model averaging (BMA) scheme (Ma et al., 2018; Rahman et al., 2020) and the Kriging-based algorithms (Manz et al., 2016; Chen et al., 2020a). These merging methods have proved to be effective in improving the accuracy of precipitation estimates. Nevertheless, most of them are subjected to strong (ad hoc) assumptions that may not hold true in reality (Wu et al., 2020). More specifically, (i) the linear average method assumes a linear relationship between the output and the merging members; (ii) the optimal interpolation technique assumes a sole dependence of the error variance on precipitation intensity; (iii) the geographically weighted regression method hypothesizes a non-linear correlation between the dependent variables; (iv) the bias correction method assumes an unbiased estimation of the precipitation background field by the SPPs; and (v) the BMA and Kriging-based methods assume a normal (Gaussian) distribution of precipitation. Moreover, many of the methods are applicable to blending gauge-based observations with only one single SPP. As we know, each SPP could have its own pros and cons, and might show varying performance for different regions and time-scales. For instance, Zhang et al. (2020a) found that the Integrated Multi-satellite Retrievals for Global Precipitation Measurement (IMERG) Early Run product outperforms the SM2RAIN-ASCAT product in the humid subregions of China, while the contrary is the case in the semi-

arid subregions. The fusion of gauge-based observations with multiple SPPs, instead of one single SPP, might be more favorable for the improvement of the precipitation estimates (Chen et al., 2020b). Lastly, to our best knowledge, there are few studies devoted to compare the performance of the merged precipitation products with the gauge-only interpolated product, especially in the conditions of different rain gauge densities. These few exceptions include the studies of Wang and Lin (2015) and Bai et al. (2019).A comparison to the gauge-only interpolated precipitation, however, is very important for the evaluation of the added value of the merged products, and could provide valuable information to the data users. For instance, the gauge-only interpolated precipitation might be more preferred than the SPPs or the merged precipitation products for hydrological modeling, providing it shows a higher performance (Zhang et al., 2020b).

The machine learning (ML) algorithms, known as data-driven methods, have received increasing popularity in hydrology and climate sciences (Sharifi et al., 2019; Wang et al., 2019; Boucher et al., 2020; Lin et al., 2020; Merz et al., 2020). Many studies have been carried out to bias-correct or merge SPPs by taking the advantages of the ML techniques. Yang and Luo (2014) corrected the Tropical Rainfall Measuring Mission (TRMM) Multisatellite Precipitation Analysis (TMPA) product by using a back-propagation (BP) neural network. Bhuiyan et al. (2018) and Ehsan Bhuiyan et al. (2019) investigated the use of the quantile regression forests (QRF) for blending multiple SPPs with the atmospheric reanalysis precipitation dataset, and evaluated its performance over the tropical regions with complex terrains. Kumar et al. (2019) corrected the near-real-time TMPA product by combining it with NRT soil moisture through a nonlinear support vector machine-based regression (SVR) model. Recently, Baez-Villanueva et al. (2020) proposed a RF-based approach to combine multiple SPPs with the gauge observations in Chile. Wu et al. (2020) presented a spatiotemporal deep fusion model to merge satellite and gauge-based precipitation in China. Bhuiyan et al. (2020) employed two ML algorithms (i.e., the RF and neural networks) to generate an error-corrected IMERG product over the Brahmaputra river basin. Wehbe et al. (2020) adopted the artificial neural networks (ANNs) to derived a multi-source precipitation product by fusing Weather Radar, Satellite Retrievals, and Surface Parameters. Previous studies demonstrated a great potential of the ML algorithms in correcting and merging SPPs. The ML technique has at least the following advantages relative to the traditional precipitation merging methods (Hengl et al., 2018; Baez-Villanueva et al., 2020; Wu et al., 2020): (i) it can deal well with the complex and non-linear relationships between inputs and output; (ii) it contains no rigid assumption; (iii) it is of great flexibility to incorporate multiple types of explanatory variables; (iv) it has the capability of blending gauge observations with multiple SPPs; and (v) it is easy to implement. Nevertheless, few of the previous studies incorporated the spatial autocorrelation information of the gauge-based observations into the precipitation merging framework. Moreover, typically, the regression model of the ML algorithm was solely used for the precipitation fusion. The combined use of the classification and regression models of the ML algorithms, which could improve the rain/no-rain detection skill (Chen et al., 2010), however, has rarely been reported in the previous studies. In addition, to our best knowledge, few efforts were made to compare the performance of different ML algorithms in merging satellite and gauge-based precipitation, and to assess the added value of their merged products relative to the gauge-only product, especially with the consideration of the influence of rain gauge density.

In the study, we proposed a novel double machine learning (DML) approach developed based on the classification and regression models of the ML algorithms to merge multiple SPPs and gauge observations. The reliability and validity of the new precipitation merging method were tested over the Chinese mainland. Our major objectives are threefold: (i) to evaluate the effectiveness of the DML algorithms; (ii) to compare the performance of DML to the traditional single machine learning (SML) methods, the linear merging methods, and the gauge-only interpolation;

and (iii) to assess the added value of the merged products of DML relative to the original SPPs and the gauge-only interpolated product. The rest of the paper is organized as follows. After the introduction, the study region and data are briefly introduced in Section 2. Section 3 describes the methodology of the study. The results and discussion are presented in Section 4 and Section 5, respectively. The conclusions are drawn in the last section.

2. Study region and data

2.1. Study region

The Chinese mainland, which covers a large geographical area of about 9.6 million km^2, was selected as the study region in this research. As depicted in Fig. 1, the Chinese mainland exhibits a strong variability of the elevations, ranging from 152 m below the sea level in eastern China to 7528 m above the sea level on the Qinghai-Tibetan Plateau (QTP). According to the gradients of elevations and the precipitation patterns, eight subregions can be subdivided from the Chinese mainland (Chen et al., 2013; Chen and Li, 2016), including Xinjiang (XJ), Qinghai-Tibetan Plateau (QTP), Northeast China (NEC), North China (NC), Southeast China (SEC).Northwest China (NWC), the middle and downstream areas of Yangtze River Basin (YZ), and the southwest Yungui Plateau (YGP). The subregions XJ and NWC have an arid and semi-arid climate with scarce precipitation and intensive evaporation; YZ, YGP and SEC have a tropical and sub-tropical climate; NEC and NC have a temperate monsoon climate of medium latitudes with hot and wet summer, and cold and dry winter; and QTP has a distinct plateau climate with strong radiation, low temperature and highly variable precipitation.

2.2. Data

2.2.1. Gauge observations

The precipitation observations at 697 rain gauges were used for training and testing the ML algorithms. As plotted in Fig. 1, the rain gauges are unevenly distributed over the Chinese mainland, with an obvious lower density in the subregions XJ, NWC and QTP than the remaining subregions. We collected the half daily precipitation observations for the period 2007–2011 from the China Meteorological Data Service Center (CMDSC, http://data.cma.cn/). The half daily precipitation are the accumulated data between 20:00–8:00 and 8:00–20:00 (Beijing time, UTC + 8), respectively. The dataset have been subjected to strict quality controls (Shen et al., 2010; Zhao and Yatagai, 2014) including (i) the manual inspection and correction; (ii) the extreme values' check; (iii) the spatio-temporal consistency check; and (iv) the internal consistency check (e.g. the duplicated data and incorrect units).

2.2.2. Satellite-based precipitation products

As listed in Table 1, three near-real-time SPPs including IMERG, PERSIANN (Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks) and GSMap (Global Satellite Mapping of Precipitation), and one bottom-up SPP generated from the ASCAT soil moisture product (i.e., SM2RAIN-ASCAT) were selected in this study. The spatial resolutions of these products were unified to $0.1° \times 0.1°$ by using the nearest neighboring interpolation algorithm to ensure a spatial consistency. The daily precipitation of the selected SPPs are consistently the accumulated data for the period 00:00–24:00 UTC, which, however, are different from the gauge observations that represent the accumulations between 20:00 and 20:00 (UTC + 8). In the study, we recalculated the daily precipitation observations by adding the accumulated data between 8:00 and 20:00 (UTC + 8) in the current day to that between
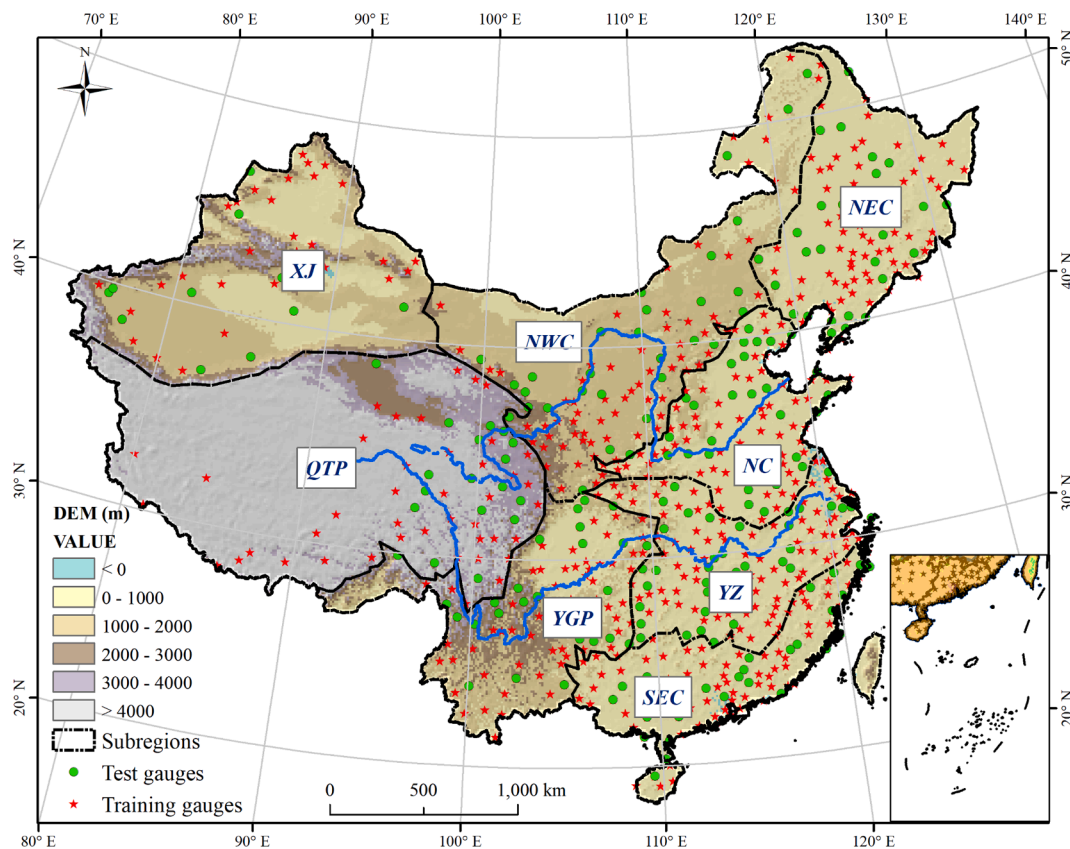

Fig. 1. Topography, the division of the subregions, and the spatial distribution of the rain gauges over the Chinese mainland. The gauges marked as red stars are the training gauges while those marked as green circles are the test gauges. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Data used in the study.

Data	Spatial and temporal resolution	Time period	Source
Gauge observations	697 gauges/half daily*	2007–2011	CMDSC http://data.cma.cn/
IMERG	0.1°×0.1°/daily	2007–2011	GES DISC https://disc. gsfc.nasa.gov/
PERSIANN	0.25°×0.25/daily	2007–2011	CHRS Data Portal http://chrsdata.eng.uci. edu/
GSMap	0.1°×0.1°/daily	2007–2011	JAXA https://hokusai.eorc. jaxa.jp
SM2RAIN-ASCAT	12.5 × 12.5 km/daily	2007–2011	Zenodo https://zenodo.org/recor d/3405563

* Half daily precipitation are the accumulated data between 20:00–8:00 and 8:00–20:00 UTC + 8, respectively.

20:00 and 8:00 (UTC + 8) in the next day, as done in our previous study (Zhang et al., 2020a). The recalculations would lead the gauge observations to be the accumulated precipitation data between 0:00 and 24:00 UTC (or 8:00 and 8:00 UTC + 8), which is in line with the SPPs.

The IMERG product was generated through the intercalibration, interpolation and integration of "all" microwave satellite-based precipitation estimates, precipitation gauge analyses, microwave-calibrated infrared (IR) satellite estimates, and other precipitation estimators (Huffman et al., 2014). It currently covers the quasi-global areas ranging 60°N-60°S with a 0.1°×0.1 spatial resolution (Huffman et al., 2019). The daily accumulated IMERG Early Run product (Version 6), simply referred to as IMERG, was used in this study. The data was downloaded from the National Aeronautics and Space Administration at the Goddard Earth Sciences Data and Information Services Center (GES DISC, https://disc.gsfc.nasa.gov/).

The PERSIANN product was produced using the network function classification/approximation procedures based on the geostationary IR brightness temperature image and daytime visible imagery (Sorooshian et al., 2000; Nguyen et al., 2019). It covers 60°S to 60°N globally with a 0.25°×0.25 spatial resolution from 2000 to present. The data, released in a near-real-time manner with 2-day delay, was developed by the Center for Hydrometeorology and Remote Sensing (CHRS) at the university of California, Irvine (UCI). We collected the data from the CHRS Data Portal at the website: http://chrsdata.eng.uci.edu/.

The GSMap product was produced using the Japan Science and Technology Agency (JAXA) Global Rainfall Watch System based on the combined MW-IR algorithm using extensive satellite data from both passive microwave (PMW) and infrared (IR) sensors (Kubota et al., 2007). It has a quasi-global coverage (60°S-60°N) with high spatial and temporal resolutions (0.1°×0.1°, 1 h). The daily accumulated GSMap Early Run product (GSMap-NRT), simply referred to as GSMap, was used in this research. The data was obtained from the Earth Observation Research Center of JAXA at the website: https://hokusai.eorc.jaxa.jp/.

The SM2RAIN-ASCAT product were generated by applying the SM2RAIN algorithm (Brocca et al., 2013) to the ASCAT soil moisture product (Brocca et al., 2019). It was provided at the global scale with a high spatial resolution of 12.5 km and a daily temporal resolution. The data was selected because it was generated form the satellite-based soil moisture product in a bottom-up manner, and could compensate the deficiency of the top-down SPPs (Zhang et al., 2020a). The newly released version of SM2RAIN-ASCAT (version 1.1), simply referred to as SM2RAIN, was collected from Zenodo at the website: https://zenodo.org/record/3405563.
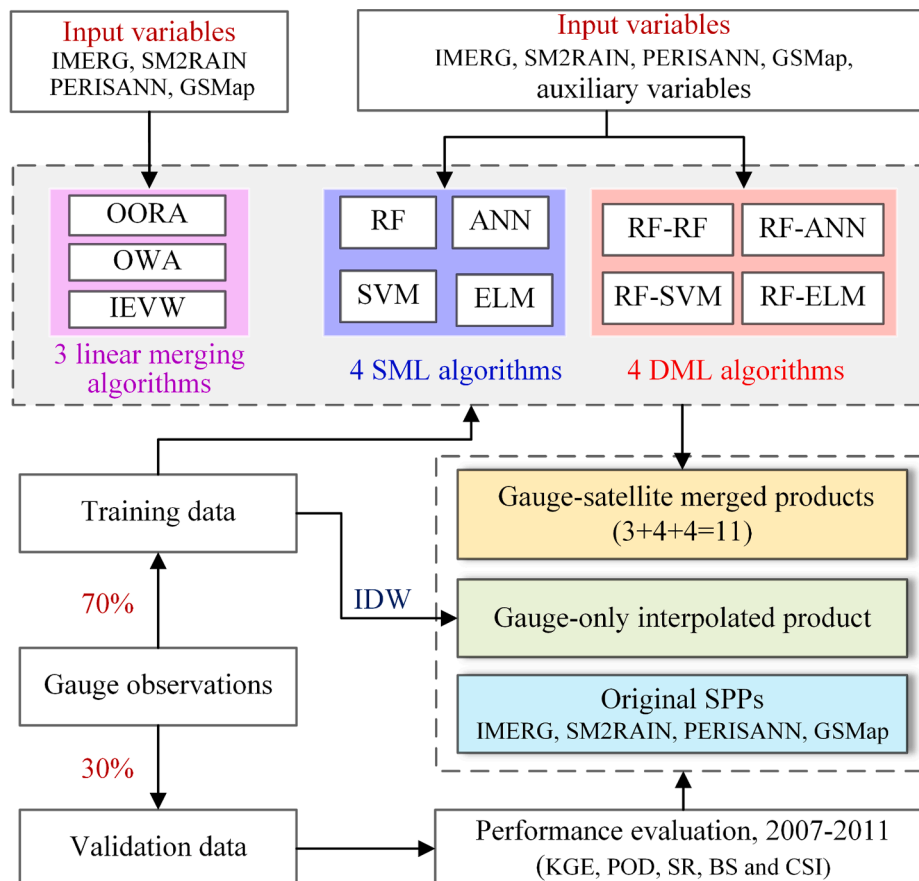

Fig. 2. Workflow of the study.

3. Methodology

The workflow of the study is briefly illustrated in Fig. 2. We adopted four traditional SML schemes and four newly proposed DML algorithms to merge multiple SPPs (i.e., IMERG, SM2RAIN, PERISANN and GSMap) and gauge observations. The SML algorithms were developed based solely on the regression models of random forest (RF), artificial neural network (ANN), support vector machine (SVM) and extreme learning machine (ELM). The DML algorithms, however, were developed based on the classification model of the random forest (RF) in combination with the regression models of RF, ANN, SVM and ELM. This leads to four different DML algorithms including RF-RF, RF-ANN, RF-SVM and RF-ELM. Moreover, the linear satellite-gauge merging methods including the inverse error variance weighting (IEVW), the one-outlier-removed average (OORA), and the optimized weight average (OWA) were also adopted in the study for the comparison with the ML-based precipitation methods. Details about the SML and DML algorithms will be presented in Sections 3.1.1 and 3.1.2, respectively, and those about the liner merging methods in Section 3.2. In addition, a gauge-only precipitation product was generated by using the simple but robust IDW (i.e. inverse distance weighting) interpolation algorithm (Camera et al., 2014). It was used together with the original SPPs as the benchmarks for the evaluation of the added value of the merged products of DML.

In the study, the gauge observations were randomly and spatially subdivided into two parts (i.e., 70% and 30%), one was used as the training dataset while the other as the test dataset. This is done by randomly sampling 70% of all the available gauges in each subregion separately, considering the uneven distribution of the rain gauges. If the sampling was performed for the entire Chinese mainland, the training and test dataset would be biased toward the southeastern regions, due to the higher rain gauge densities. In total, this study produced twelve precipitation products including four of the DML algorithms, four of the SML algorithms, three of the linear merging methods, and one additional gauge-only interpolated product. These products and the original SPPs were evaluated based on the test dataset by using both the continuous and categorical metrics including the Kling-Gupta efficiency (KGE), probability of detection (POD), success ratio (SR), bias score (BS) and critical success index (CSI).

3.1. Machine learning-based satellite-gauge merging methods

Fig. 3 presents the framework of the ML-based methods used to merge SPPs and gauge observations. A critical and preliminary step is to choose informative auxiliary variables. The geographical and topographical covariates including latitude, longitude, subregion ID and altitude were selected to account for the spatial variations of precipitation and the influences of topography on precipitation patterns. The other topographical variables including slope, aspect, terrain shadows and roughness were tested to have minimal impacts on the merged results, and they were excluded from the ML models. In addition, the geographical correlation information (i.e., the spatial autocorrelation) of the precipitation observations were incorporated into the
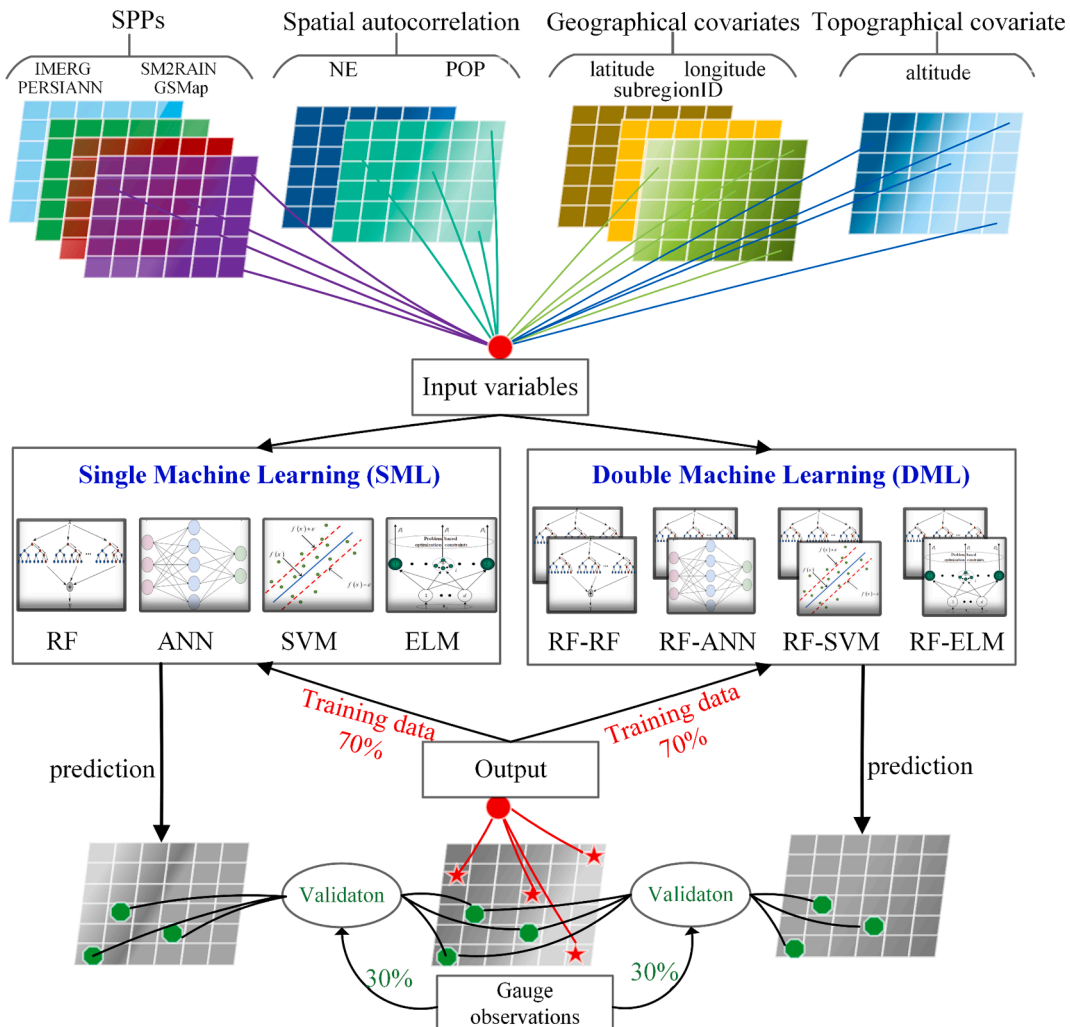

Fig. 3. Framework of the machine learning-based methods used to merge multiple SPPs and gauge observations.

precipitation merging framework, following the previous studies (Li et al., 2017; Hengl et al., 2018; Baez-Villanueva et al., 2020). The rationale behind this is that the nearer observations are more informative than the farther ones for the precipitation estimates. We defined two covariates (i.e., *NE* and *POP*) to account for the spatial autocorrelation between the gauge observations.

$$NE = \frac{\sum_{i=1}^{n} w_i P_i}{\sum_{i=1}^{n} w_i}, w_i = \frac{1}{d_i^2} \tag{1}$$

$$POP = \frac{\sum_{i=1}^{n} w_i PO_i}{\sum_{i=1}^{n} w_i}, w_i = \frac{1}{d_i^2}, PO_i = \begin{cases} 0, P_i = 0 \\ 1, P_i > 0 \end{cases} \tag{2}$$

where *NE* is the precipitation estimates from the neighbor observations; d_i is the distance to the ith neighbor gauge; *POP* is the probability of precipitation occurrence (Thornton et al., 1997); PO_i is the binary variable related to occurrence of observed precipitation; and P_i is the precipitation at the ith neighbor gauge.

After the determination of the auxiliary variables, the SML (Section 3.1.1) and DML (Section 3.1.2) algorithms were trained and calibrated at each day of the period 2007–2011 by using the training dataset (i.e. the observations at 493 gauges, Fig. 1). The day-to-day training strategy was tested to be more effective and computationally efficient than the periodic training strategy (e.g., month-to-month and year-to-year). Afterwards, the trained ML algorithms were used to predict daily precipitation at the test gauges for the same period (i.e., 2007–2011). The observations at the test gauges, which are completely independent from the training dataset, would benefit a robust performance evaluation of the ML algorithms. The assessments were carried out at the daily scale over the entire study period (2007–2011), with some attention paid to the seasonal and subregional variations of the performance.

3.1.1. Single machine learning approach

The principle of the SML-based merging method is to build a transfer function between the predictors and the output based on the regression model of the ML algorithm as:

$$P_{merg} = f(X_s, X_a) + error \tag{3}$$

where P_{merg} is the output, which represents the gauge-based and merged precipitation, respectively, at the training and test/prediction stages; *f* is the transfer function built with the ML regression model; X_s and X_a are the predictors, of which the former represents the precipitation estimates of SPPs while the later represents the auxiliary variables. In the study, the popular ML algorithms including RF, ANN, SVM and ELM were used to construct the transfer functions, which leads to four different SML-based precipitation merging methods.

(1) Random Forest (RF)

The RF algorithm is an ensemble ML approach that aggregates the results of multiple decision tree models (Breiman, 2001). The randomness of RF mainly reflects in two aspects: one is the random samples, and the other is the random features or predictors. The bootstrap method is used to generate the random samples from the training dataset, which are further adopted to construct a number of decision trees. At each decision tree, the internal splits are generated according to the randomly selected features. RF could substantially alleviate the overfitting problems and improve the generalization capability, owing to the random selections of the input samples and predictors, and the use of ensemble predictions. This study implemented the RF algorithm using the MATLAB TreeBagger function. As shown in Table 2, the hyperparameters including the *Ntree* and *MinObs* were determined for the RF algorithm through the trial and error procedures. They were identified as 50 and 5, respectively. The hyperparameter *Nsplit*, i.e., the number of variables randomly sampled at each decision split, was set to one third of the number of the input variables, following the previous studies (Beck

Table 2
Descriptions and calibration methods of the hyperparameters for the different ML algorithms.

ML algorithm	Hyperparameter	Descriptions	Calibration method
RF	*Ntree*	Number of trees	Trial and error
	MinObs	Minimum number of observations per node	Trial and error
ANN	*NNH*	Numbers of nodes in the hidden layer	Parallel computing-based and grid-research (PC-GR)
	IW	Initial weight	PC-GR
SVM	*C*	Penalty parameter	PC-GR
	γ	Parameter of the radial basis function (RBF)	PC-GR

et al., 2019; Baez-Villanueva-test et al., 2020).

The TreeBagger has a built-in function to measure the relative importance of the different predictors. This is done through the method based on Permutation Feature Importance (PFI) by using the out-of-bag (OOB) samples (i.e., the samples not selected by the bootstrap approach) (He et al., 2016; Schmidt et al., 2020). The underlying principle of PFI is to broke the ties between the target and one of the predictors, and then estimate the prediction accuracy deterioration (Schmidt et al., 2020). The stronger of the deterioration, the more important of the predictor for the model's prediction. More specifically, PFI is estimated with the following steps: (i) calculating the original prediction error (i.e., the mean squared error for the regression tree, and the misclassification probability for the classification tree) across the OOB observations for the jth decision tree (i.e., $ErrOOB_j$); (ii) generating the permuted predictor matrix by shuffling the ith predictor while keeping the others fixed, and re-estimating the prediction error for the jth decision tree (i.e., $ErrOOB_{i, j}$); (iii) calculating the increase in the prediction error ($IncreError_{i, j}$) after shuffling the ith predictor using the Eq.4; and lastly, (iv) estimating PFI of the ith predictor (PFI_i) by averaging the $IncreError_{i, j}$ and dividing it by the standard deviation (σ) over the decision trees using Eq.5.

$$IncreError_{i,j} = ErrOOBi_{i,j} - ErrOOB_j \tag{4}$$

$$PFI_i = \frac{\frac{1}{Ntree}\sum_j^{Ntree} IncreError_{i,j}}{\sigma(IncreError_{i,j})}, i = 1, 2, 3 \cdots 10 \tag{5}$$

(2) Artificial Neural Network (ANN)

The ANN algorithm is an information processing paradigm inspired by the biological neural networks (Gardner and Dorling, 1998). The basic elements of ANN are the neurons (or units), which are interconnected by the weighted links. At each unit, the output is computed by a transfer (or active) function of the weighted summation of the inputs. The multilayer perceptron (MLP) with a three-layer structure (i.e., the input, hidden and output layers), which are one of the widely used forms of the ANN algorithms, was used in the study. We implemented the ANN algorithm using the MATLAB fitnet function. This study setted the transfer functions (*TranFunc*) as 'tansig' for both the hidden and output layers of ANN. As presented in Table 2, the hyperparameters including the *NNH* and *IW* were optimized through a parallel computing based and grid-research (PC-GR) approach. More specifically, we considered 3 *NNH* (i.e., 10,20, 30) and 10 random *IW* (i.e., 30 grids), and searched the best 10 grids through the cross validation. The searching processes were implemented via the parallel computing technique to improve the computational efficiency. The ensemble average of the precipitation estimations with the best 10 parameter sets were used as the final predictions.

(3) Support Vector Machine (SVM)

The SVM is a popular ML algorithm widely used for both classification and regression purposes (Vapnik, 1998). The classification model of the SVM aims to find a hyperplane whereby the separate categories are divided by a clear gap that is as wide as possible. On the other hand, the regression model aims to find a hyperplane expressed by a regression function that best describes the observed outputs with an error tolerance. In addition to linear problems, SVMs can also deal efficiently with non-linear problems by using the kernel function to map the non-linear inputs to linearly separable and high-dimensional feature spaces. In this study, the SVM algorithm was implemented by using the LibSVM package (Chang and Lin, 2011). The kernel function (*KernFunc*) was set as the radial basis function (RBF) through the trial and error procedure. The two-stage PC-GR approach was applied to optimize the two hyperparameters of SVM including C and γ. We first determined the best parameter set ($C1$ and $\gamma1$) from the coarse grids ($C = \gamma = -8\text{1}0{:}2{:}10$) through the cross validation, and then searched the fine grids centered on $C1$ and $\gamma1$, i.e., ($C1$-2):0.2:($C1 + 2$), ($\gamma1$-2):0.2:($\gamma1 + 2$), to identify the final best parameter set. Similarly, the parallel computing technique was used to accelerate the optimization and training processes.

(4) Extreme Learning Machine (ELM)

The ELM algorithm proposed by Huang et al. (2012) has the similar structure as the MLP with one input layer, one hidden layer and one output layer. After randomly initializing the weights and bias between the input and hidden layers, the weights between the hidden and output layers are directly estimated as the product of the Moore-Penrose generalized inverse of the output matrix of the hidden layer and the target variable. This is quite different from the MLP that estimates the weights between the hidden and output layers via an iterative learning process. The distinct feature leads to a high computational efficiency of the ELM algorithm. The codes of ELM are available at: https://www.ntu.edu.sg/home/egbhuang/elm_codes.html. We modified the original codes to get an ensemble prediction of the 10 best estimates of ELM through the parallel computing-based optimization process similar to ANN. These efforts are mainly devoted to ameliorate the impact of the random initial weights and biases between the input and hidden layers, and the influence of different number of nodes in the hidden layer.

3.1.2. Double machine learning approach

Inspired by the work of Chen et al. (2010), this study proposed a novel DML approach based the classification and regression models of the ML algorithm to merge multiple SPPs and gauge observations. As presented in Fig. 4, we first developed a classification model by using one of the ML algorithms (i.e., RF, ANN, SVM and ELM). The precipitation observations that are greater than zero were classified as 'wet day', while those are equal to zero as 'dry day'. The classified precipitation observations were used together with the input variables to train and optimize the classification model of each ML algorithm, which refers to the first ML. The training and optimization processes are similar to the regression models of the ML algorithms (Section 3.1.1), with the difference that the optimization objectives are the classification accuracy of the precipitation events. Afterwards, the classification accuracy of 'dry/wet day' for each ML algorithm was evaluated based on the independent test dataset (Fig. 1); and the best performed ML algorithm was identified. The classification accuracy (*CA*) was defined as:
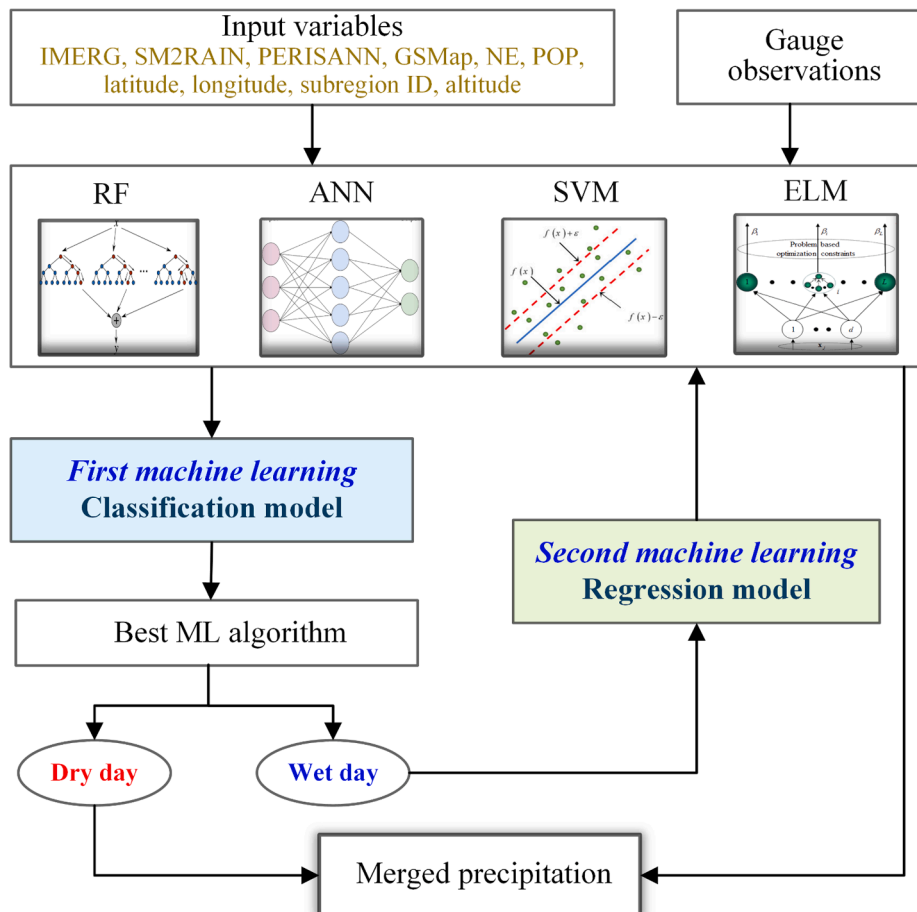

Fig. 4. Schematic of the double machine learning (DML) algorithm.

$$CA = \frac{1}{N} \sum_{i=1}^{N} \frac{W_i + D_i}{T_i} \qquad (6)$$

where W_i and D_i are the total number of days correctly classified as 'wet day' and 'dry day', respectively, at the ith gauge; T_i is the total number of days over the evaluation period (i.e., 2007–2011); and N is the total number of gauges in the test dataset (i.e., 204 in our case). At the second stage, we trained the regression model of each ML algorithm to estimate precipitation at the day that was predicted as 'wet day', similar to the SML approach, which refers to the second ML. The amount of precipitation was set to zero if the day was predicted as 'dry day' by the classification model of the ML algorithm. Hence, the merged products of DML were generated by combining the predictions of the classification and regression models of the ML algorithms at the two separate stages.

The 'wet/dry day' classification accuracy of the classification and regression models of the ML algorithms (i.e., RF, ANN, SVM and ELM) are listed in Table 3. The classification models show very comparable classification accuracy (0.888–0.889), which are obviously higher than the regression models (0.286–0.450). The preliminary analysis indicates that the inclusion a classification model into the precipitation merging framework is reasonable, and could improve the rain/no-rain detection skill. In this study, we selected the classification model of RF for the classification purpose, due to its slightly better performance and higher computational efficiency relative to the other ML algorithms. This led to four different DML algorithms including RF-RF, RF-ANN, RF-SVM and RF-ELM.

3.2. Linear satellite-gauge merging methods

This study adopted three linear precipitation merging methods, including the inverse error variance weighting (IEVW), the optimized weight average (OWA), and the one-outlier-removed average (OORA), which are described in Eqs. (7), (8) and (9), respectively.

$$P_{merg} = \frac{1}{\sum_{i=1}^{N} 1/e_i^2} \sum_{i=1}^{N} \frac{1}{e_i^2} \times S_i \qquad (7)$$

$$P_{merg} = \sum_{i=1}^{N} wopt_i \times S_i \qquad (8)$$

$$P_{merg} = \frac{1}{N-1} \sum_{i=1}^{N-1} S_i \qquad (9)$$

where P_{merg} is the merged precipitation; N is the number of SPPs; S_i is the ith SPP; e is the error variance; and $wopt$ is the optimized weight. The error variance and weight of each SPP were calculated for the IEVW method based on the training dataset; and the weights of SPPs were optimized for the OWA method with the objective of minimizing the root mean square error (RMSE) between the precipitation estimates and the observations in the training dataset. Afterwards, the weights were used by the IEVW and OWA methods to predict precipitation at the test gauges. Regarding to the OORA method, one of the SPPs that has the largest RMSE was removed and regarded as the outlier, and the arithmetic average of the precipitation estimates of the remaining SPPs was used as the merged results.

Table 3
Wet/dry day classification accuracy of the classification and regression models of the RF, ANN, SVM and ELM algorithms.

Model	Machine learning (ML) algorithms			
	RF	ANN	SVM	ELM
Classification	0.899	0.895	0.894	0.888
Regression	0.450	0.296	0.493	0.573

3.3. Performance evaluation

The Kling-Gupta efficiency (KGE) proposed by Gupta et al. (2009) was used for the performance evaluations. As shown in Eq.10, the KGE is a multi-component performance metric that includes the correlation coefficient (CC), the bias ratio (Beta), and the variability ratio (Gama), which measure the correlation, bias and relative temporal variability between the precipitation estimates and the gauge observations, respectively.

$$KGE = 1 - \sqrt{(1 - CC)^2 + (1 - Beta)^2 + (1 - Gama)^2} \qquad (10)$$

$$CC = \frac{\sum_{i=1}^{n}(P_i^{merg} - P_{mean}^{merg})(O_i - O_{mean})}{\sqrt{\sum_{i=1}^{n}(P_i^{merg} - P_{mean}^{merg})^2} \sqrt{\sum_{i=1}^{n}(O_i - O_{mean})^2}} \qquad (11)$$

$$Beta = \frac{\mu_{merg}}{\mu_O} \qquad (12)$$

$$Gama = \frac{\sigma_{merg}}{\sigma_O} \qquad (13)$$

where P and O are the merged precipitation and the observations, respectively, at the test gauges; μ is the mean value; σ is the standard deviation; i is the time step; and n is the total number of time steps. In the study, following Kling et al. (2012), we calculated the Gama by using Eq. (14), instead of E q.13, to ensure the variability and bias ratios are not cross-correlated.

$$Gama = \frac{\sigma_{merg}/\mu_{merg}}{\sigma_O/\mu_O} \qquad (14)$$

The ranges of the KGE, CC, Beta and Gama are $[-\infty, 1]$, $[-1, 1]$, $[0, \infty]$ and $[0, \infty]$, respectively. The closer of the KGE and its components to unity, the better the performance of the precipitation products.

In addition, we adopted four categorical metrics to assess the capability of the products in detecting precipitation events. They are the probability of detection (POD), success ratio (SR), bias score (BS) and critical success index (CSI), which are defined as in Eqs. (15), (16), (17) and (18), respectively.

$$POD = a/(a + c) \qquad (15)$$

$$SR = a/(a + b) \qquad (16)$$

$$BS = \frac{POD}{SR} = (a + b)/(a + c) \qquad (17)$$

$$CSI = a/(a + b + c) \qquad (18)$$

where a is the number of precipitation events correctly detected; b is the number of non-events that are incorrectly detected (i.e., event was detected to occur but not observed to occur); c is the number of the missing events (i.e., event was not detected to occur but observed to occur). The four metrics were mapped on a single diagram proposed by Roebber (2009) to better visualize the evaluation results. The closer of the metrics to the upper right of the diagram, the better detectability of precipitation events. The perfect performance is reached when POD = SR = CSI = BS = 1. The categorical metrics were computed at each test gauge for the precipitation thresholds of 1, 5, 10 and 15 mm/day, respectively.

3.4. Sensitivity analysis of training dataset size

A sensitivity analysis was additionally conducted to investigate the impacts of training dataset size on the performance of the DML algorithms and the gauge-only interpolation. This is done by reducing the size of the original training dataset, and analyzing the changes of the

accuracy of the precipitation estimates at the test gauges. Specifically speaking, the size of the training dataset (i.e., the proportions of rain gauges in the training dataset) was first randomly reduced to 80%, 60%, 40% and 20% of its original size that includes 493 gauges. The reduction processes were implemented for each subregion separately to ensure the condensed training dataset is representative over the Chinese mainland. Afterwards, we repeated the DML-based merging process and the interpolation process with the different sizes of the training dataset. Lastly, we analyzed the variations of the performance of the DML algorithms and the gauge-only interpolation with the decreasing sizes of the training dataset. The sensitivity analysis would facilitate an evaluation of the robustness and reliability of the DML algorithms.

4. Results

4.1. Performance assessment over the Chinese mainland

4.1.1. Overall performance

Fig. 5 shows the boxplots of the KGE and its components (CC, Beta and Gama) for the sixteen precipitation products over the Chinese mainland. The median KGE values are relatively small for GSMap, PERISANN and SM2RAIN, which are 0.39, 0.31 and 0.32, respectively (Fig. 5a). The IMERG product performs the best among the original

SPPs, with a median KGE of 0.54. The merged products produced by the three SML algorithms (i.e., RF, ANN and ELM) achieve higher median KGE values than the original SPPs, which are 0.65, 0.60 and 0.60, respectively. The median KGE of the merged product generated by SVM, however, is smaller than the IMERG product. The products of the linear merging methods (i.e., OORA, OWA and IEVW) show comparable performances to the IMERG product in terms of KGE. Regarding to the gauge-only interpolated product (i.e., IDW), interestingly, it provides a comparable and even higher median KGE than the products of the SML algorithms and the linear merging methods. The merged products of the DML algorithms (i.e., RF-RF, RF-ANN, RF-SVM and RF-ELM) have consistently higher median KGE (0.67–0.71) than the other precipitation products. Fig. 5b shows the boxplots of the correlation coefficients (CC) for the different precipitation products. The merged products of SML and DML give comparable median CC, which are higher than other products. The original SPPs have the lowest CC (ranging 0.46–0.64) among the sixteen precipitation products. The median bias ratios (Beta), as show in Fig. 5c, are closer to unity with smaller ranges for the DML-based and RF-based products, and the gauge-only interpolated product, in comparison to other products. As present in Fig. 5d, the median variability ratio (Gama) for the products of DML range from 0.83 to 0.90, which are apparently higher than those of SML (0.66–0.76), and the linear merged and gauge-only interpolated products (0.70–0.72). The median Gama of
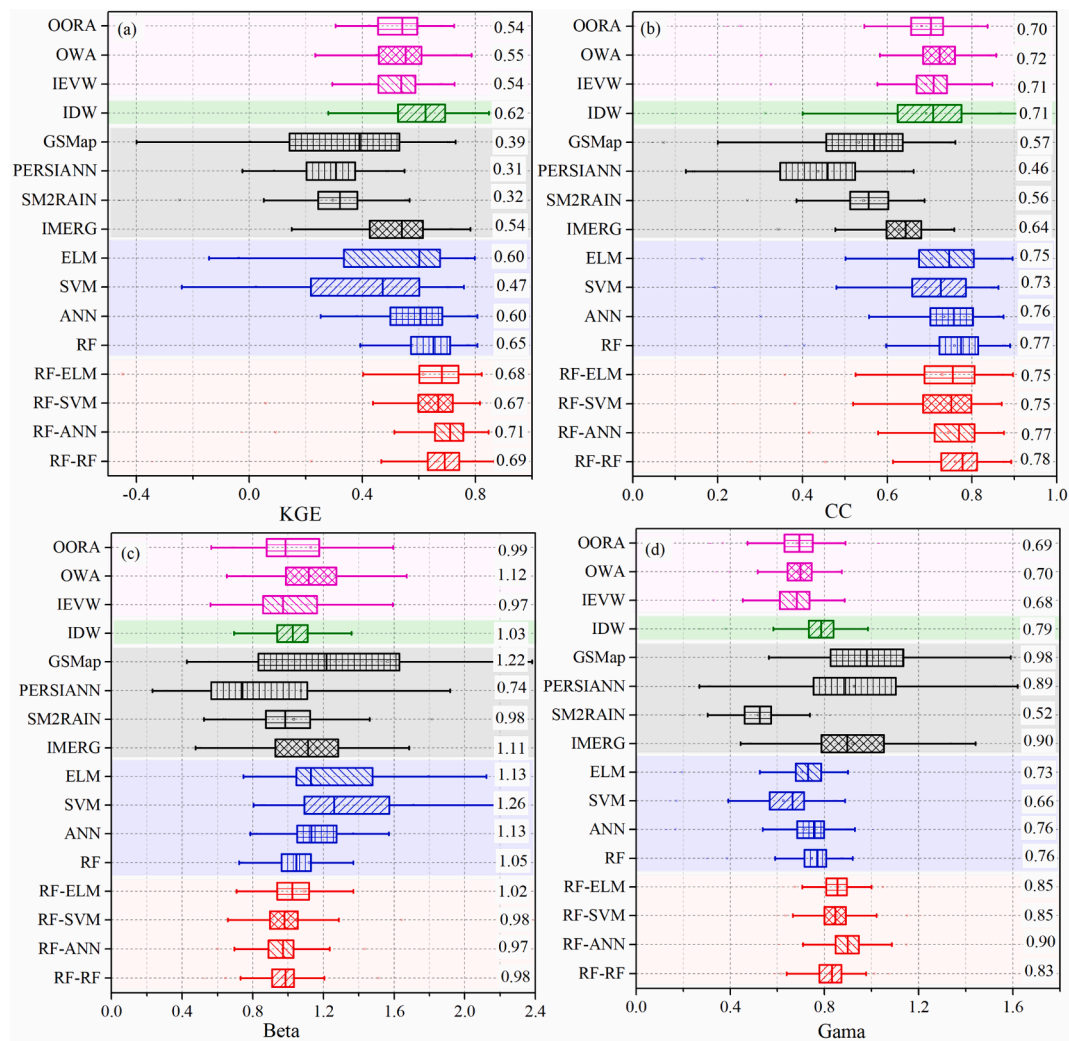

Fig. 5. Boxplots of the KGE and its components (CC, Beta and Gama) for the merged precipitation products of the DML (red color) and SML (blue color) algorithms, the original SPPs (black color), the gauge-only interpolated product (green color) and the linear merged products (purple color) over the Chinese mainland. The median KGE values are provided at the top of each box. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

IMERG, GSMap and PERISANN are close to unity, but with large ranges in comparison to the other products. Overall, the merged products of the DML algorithms outperform the other merged products and the gauge-only interpolated product, in terms of KGE.

Fig. 6 shows the Roebber's performance diagrams for the sixteen precipitation products with different precipitation thresholds over the Chinese mainland. The merged products of the DML algorithms exhibit the best performance in detecting precipitation events with the threshold of 1 mm/day, followed by the RF-based product and the gauge-only interpolated product. The linear merged products and the products of ANN, SVM and ELM achieve comparable median CSI values but larger BS to the IMERG and GSMap products. With respect to the precipitation events higher than 5 mm/day (i.e., precipitation threshold $\geq$ 5 mm/day), the merged products of the DML algorithms perform comparable to those of the SML algorithms except SVM. They consistently perform better than the other products, as shown in Fig. 6a, c and d. Comparing the results for different precipitation thresholds, all the products tend to show lower detectability of the precipitation events with larger magnitudes, as indicated by the decreasing CSI values. Regardless, the merged products of the DML algorithms outperform the other products in detecting precipitation events with the threshold of 1 mm/day, and perform better than the original SPPs regardless of the precipitation thresholds.

4.1.2. Seasonal performance

Fig. 7 presents the boxplots of the KGE and CSI for the 16 precipitation products in the seasons of spring (March-May), summer (June-August), autumn (September-November), and winter (December-February). In terms of KGE, the original SPPs consistently perform worse than the products of the DML algorithms in all of the seasons. The linear merged products obviously outperform GSMap, SM2RAIN and PERSIANN in spring, autumn and winter, while they perform worse than the gauge-only interpolated product. The IMERG product shows a comparable performance to the merged product in spring and summer, although it exhibits a lower performance in autumn and winter. The products of the SML algorithms including ELM, SVM and ANN generally perform better than the original SPPs, but they perform worse than the gauge-only interpolated product in the different seasons. The products of the DML algorithms show higher performances than other products in terms of KGE in spring, summer and autumn. In winter, the products of the DML and RF algorithms and the gauge-only interpolation achieve comparable higher performances than the other products. In terms of CSI, the products of the DML algorithms outperform all the other products in spring, summer and autumn with the precipitation threshold of 1 mm/day, followed by RF-based product and the gauge-only interpolated product, as shown in Fig. 7e, f and g. In winter, the products of the DML, RF and ANN algorithms perform comparable to the gauge-only interpolated product, but outperform the remaining products.

4.2. Performance assessment at the subregional scale

Fig. 8 shows the boxplots of the KGE for the sixteen precipitation products over the eight subregions of the Chinese mainland. The merged products of the DML algorithms perform better than the other products in most of the subregions including SEC, YZ, NC, YGP, NEC, and NWC. In the subregion QTP, the product of RF-RF performs best among the
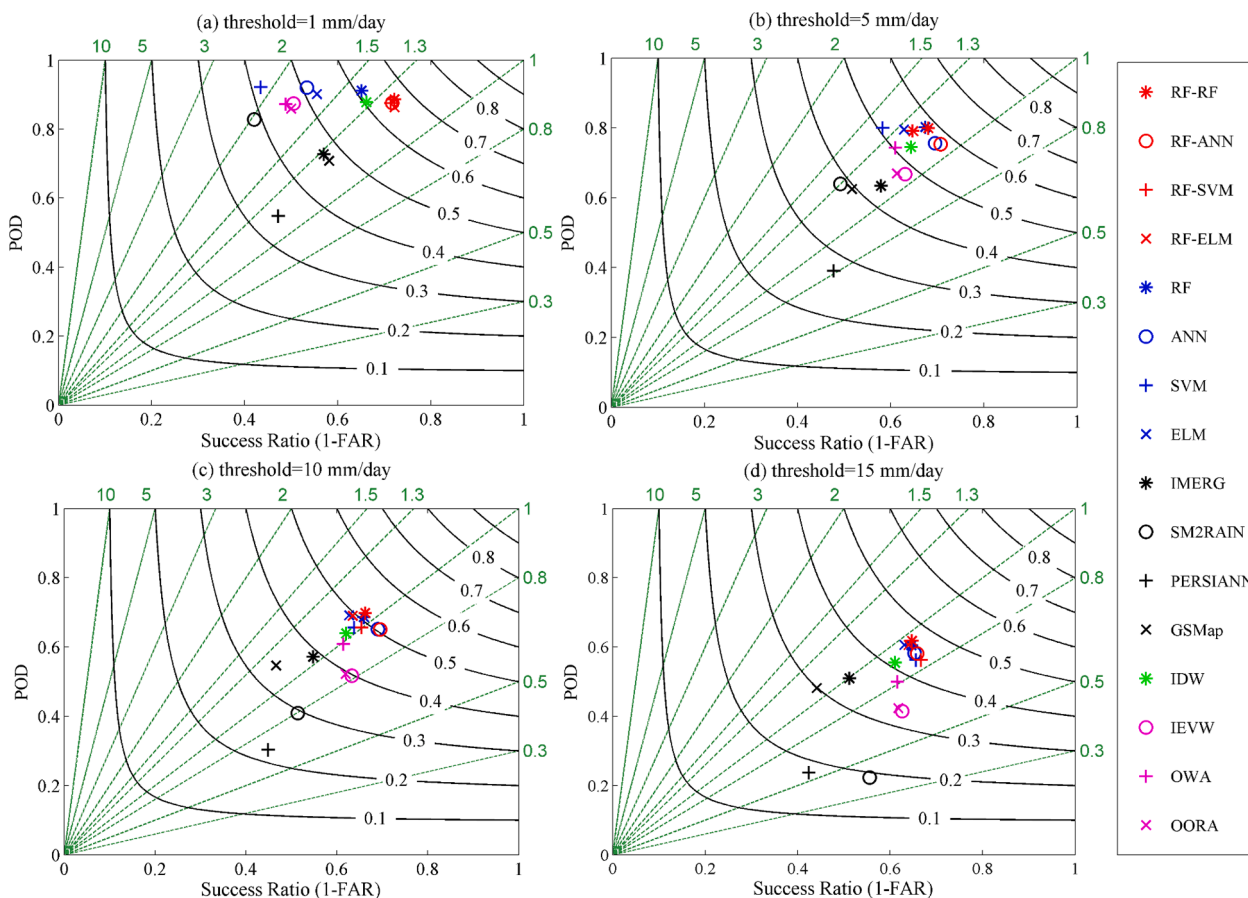

Fig. 6. Roebber's performance diagram for the merged precipitation products of the DML (red color) and SML (blue color) algorithms, the original SPPs (black color), the gauge-only interpolated product (green color) and the linear merged products (purple color). The green dash line labels on the outward extension of the line represents bias score (BS), while the black solid line represents critical success index (CSI). The four diagrams present the results for the precipitation thresholds of 1, 5, 10 and 15 mm/day, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
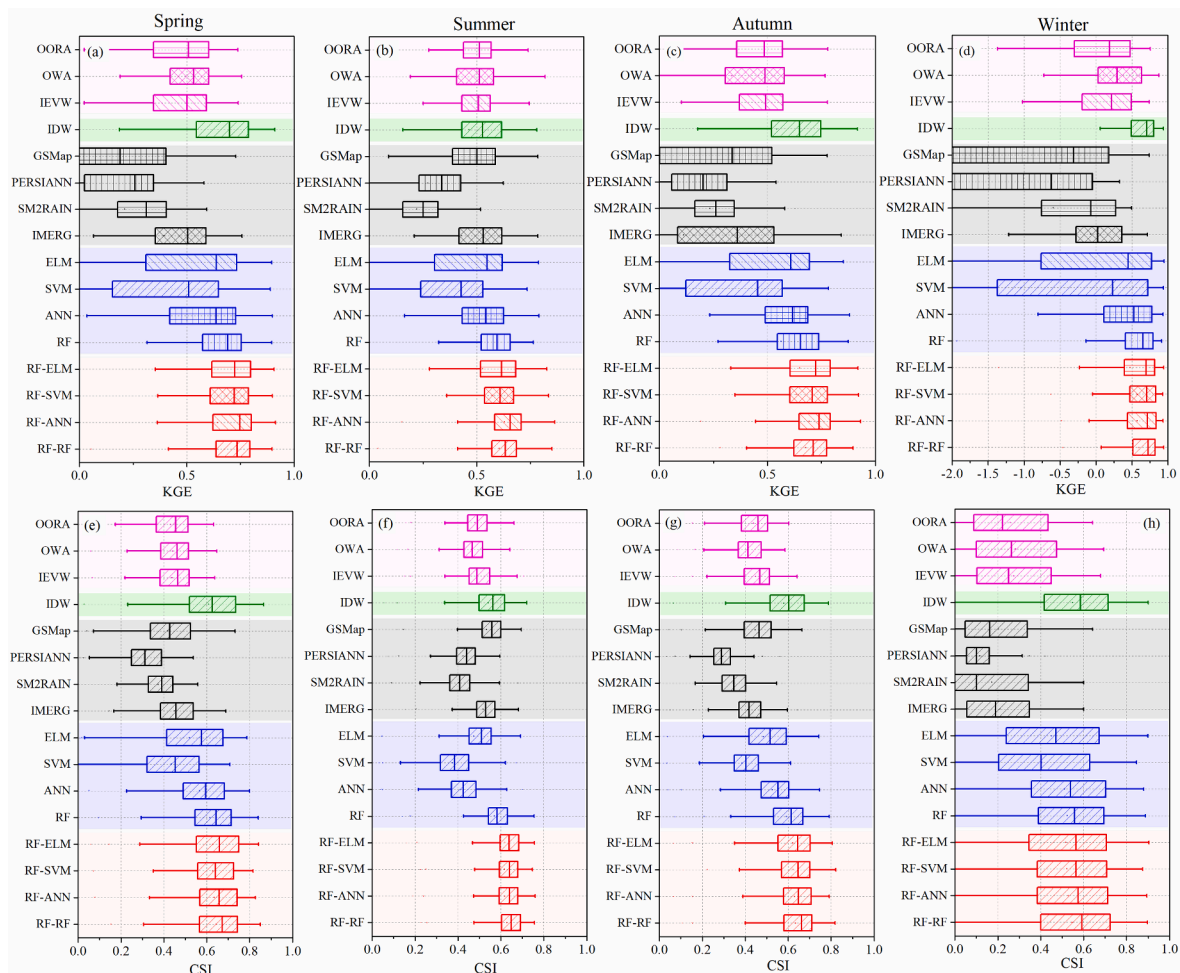
Fig. 7. Boxplots of the KGE (top panel) and CSI (bottom panel) for the merged precipitation products of the DML (red color) and SML (blue color) algorithms, the original SPPs (black color), the gauge-only interpolated product (green color) and the linear merged products (purple color) in spring, summer, autumn and winter. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

sixteen precipitation products, followed by those of RF-ANN, RF and RF-SVM. The products of RF-RF, RF-ANN and RF-SVM perform better than the other products in the subregion XJ, while the product of RF-ELM perform worse than SM2RAIN and the gauge-only interpolated product. Fig. 9 presents the Roebber's performance diagrams for the different precipitation products over the different subregions. The products of DML outperform the other products in detecting precipitation events (threshold = 1 mm/day) in all the subregions, especially in the subregion XJ. The RF-based product and the gauge-only interpolated product show higher performances than the other products except those of DML. All of the precipitation products tend to perform worse in XJ than in the other subregions. The performances for the merged and interpolated products can vary substantially over the eight subregions. This can be explained by the varying densities of the rain gauges used for the training of the ML algorithms or for the interpolation, and the varying complexities of the spatial–temporal variabilities of precipitation. Nevertheless, our results indicate that, in terms of KGE and CSI, the merged products of the DML algorithms generally perform better than the other precipitation products at the sub-regional scale.

4.3. Influence of training dataset size on the performance of the DML algorithms

Fig. 10 shows the boxplots of the KGE for the DML-based precipitation products (i.e., RF-RF, RF-ANN, RF-SVM and RF-ELM) and the gauge-only interpolated product (i.e., IDW) at the test gauges with the

different sizes of the training dataset. The KGE ranges of the best performed SPP (i.e. IMERG), which is independent of the size of the training dataset, were also provided for the comparison purpose. We can observe the decreasing trends of the performances for the merged and gauge-only interpolated products when the size of the training dataset (i.e., the proportions of rain gauges in the training dataset) reduces from 100% to 20%. Nevertheless, the deteriorations of the performance are more significant for the gauge-only interpolated product than the DML-based product. Taking a close look, we can note that the performance of the interpolated product is comparable and even worse than the best SPP when the proportion of gauges in the training dataset drops below 40%. The DML-based products, however, consistently perform better than the best SPP regardless of the size of the training dataset. Meanwhile, the performances of the DML-based products only show slight decreases when the size of the training dataset drop from 100% to 60%. The sensitive analysis implies that the newly proposed DML approach is robust, and has obvious advantages over the gauge-only interpolation in the condition of low training dataset size.

5. Discussion

5.1. Relative importance of the predictors

An analysis of the relative importance of the predictors could help to better understand and interpret the ML algorithm. In this study, the relative importance of the selected predictors (i.e., Permuted Feature
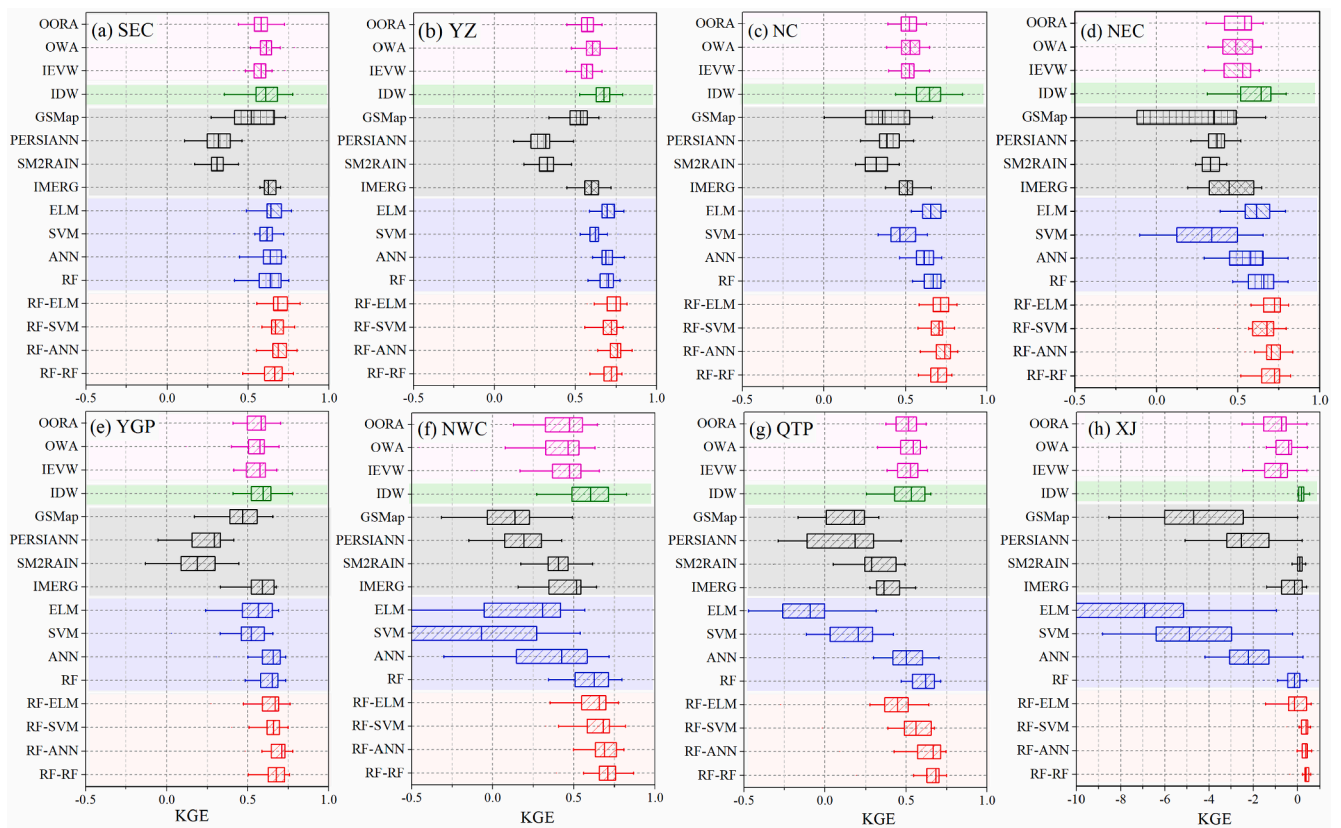
Fig. 8. Boxplots of the KGE for the sixteen precipitation products over the eight subregions of the Chinese mainland.
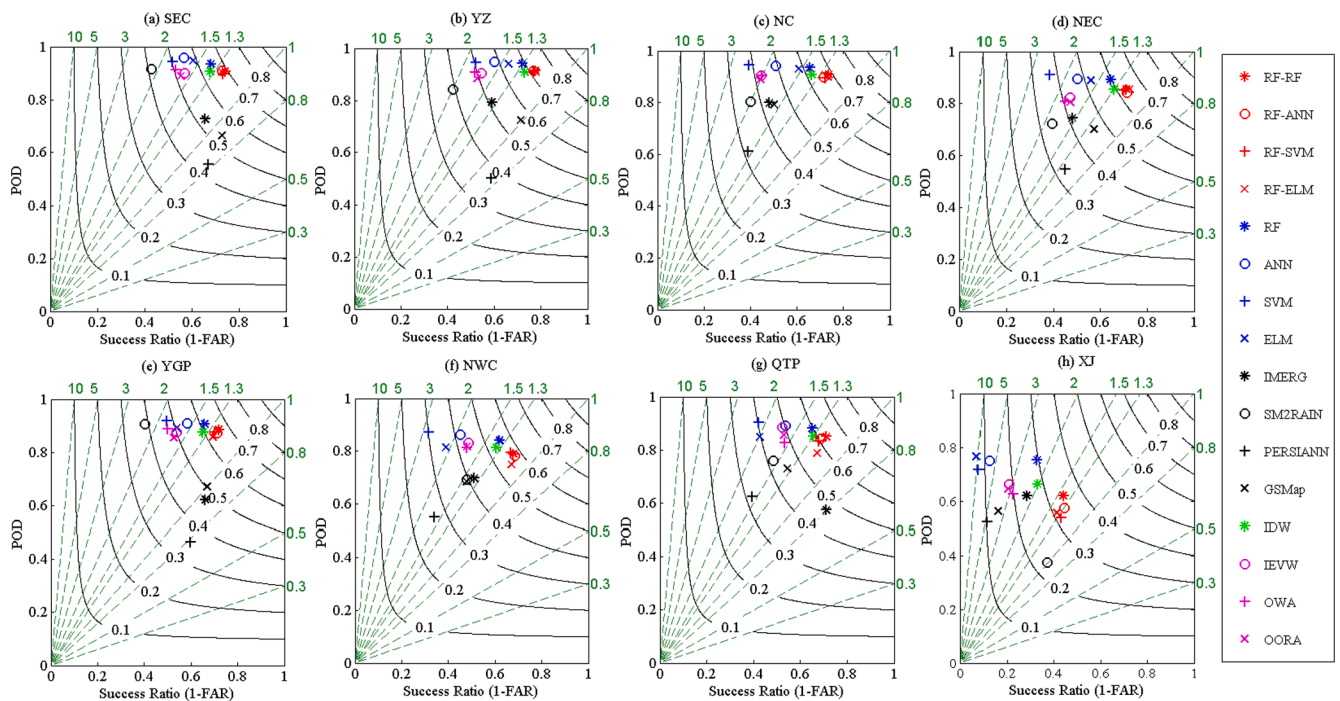

Fig. 9. Roebber's performance diagrams for the sixteen precipitation products (threshold = 1 mm/day) over the eight subregions of the Chinese mainland. The green dash line represents bias score (BS), while the black and labeled solid line represents critical success index (CSI). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Importance, PFI) for the classification and regression models of the RF algorithm, which are employed by the DML approach (i.e., RF-RF), were estimated using the MATLAB TreeBagger function. As shown in Fig. 11,

all the predictors achieve the PFI higher than 0.15 for both the classification and regression models, indicating that they could contribute useful information to the precipitation merging results. The spatial
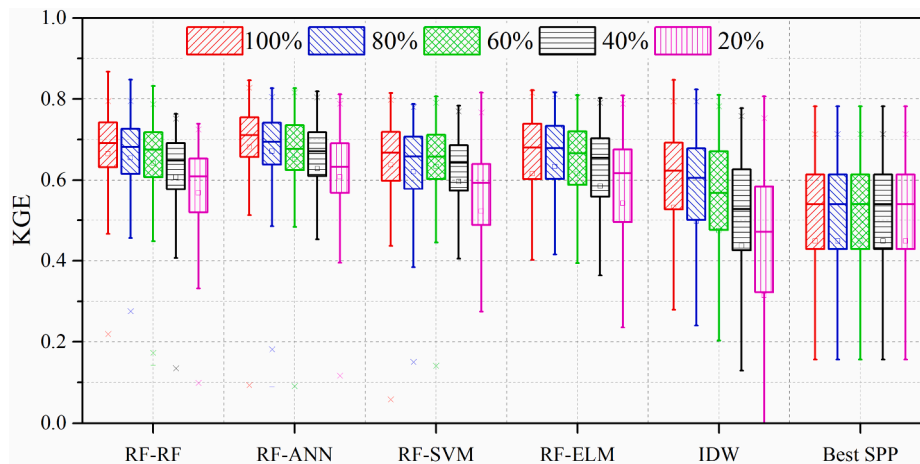
Fig. 10. Boxplots of the KGE for the DML-based precipitation products, the gauge-only interpolated product (IDW), and the best SPP with the different sizes of the training dataset.
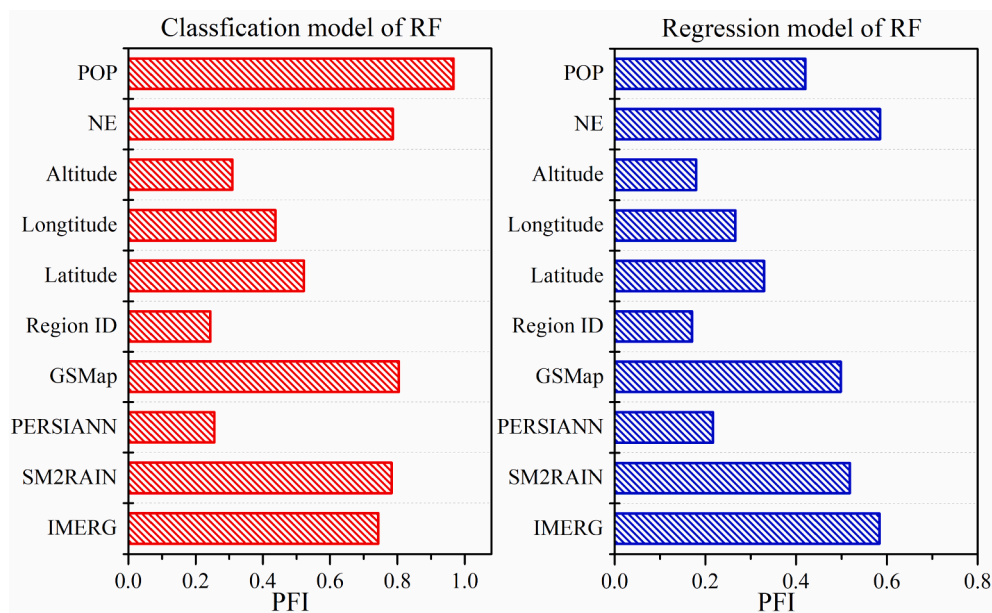

Fig. 11. Permuted Feature Importance (PFI) of the selected predictors for the classification (left) and regression (right) models of the RF algorithm.

autocorrelation predictors (i.e., *NE* and *POP*) and the SPPs including GSMap, SM2RAIN and IMERG exhibit higher importance than the other predictors. This indicates that it is very important to include the spatial autocorrelation information of the gauge observations in the ML-based precipitation merging framework, which could help to incorporate the geographical laws into ML (Yuan et al., 2020). The PERSIANN product shows an obviously lower importance than the other SPPs (i.e., GSMap, SM2RAIN and IMERG), which might be explained by its coarser spatial resolution. The time-invariant covariates including altitude, longitude, attitude and Region ID present are of relatively low importance, which is consistent with the previous reports (Bhuiyan et al., 2018, 2020; Baez-Villanueva-test et al., 2020).

5.2. Added value of the merged products of the DML algorithms

A quantitative assessment of the added value of the merged products relative to the merging members is of great importance for evaluating the effectiveness of the satellite-gauge merging method. In the study, we further analyzed the variations of the added value (i.e., the increase in KGE, ΔKGE) of the DML-based products (i.e., RF-ANN and RF-RF)

relative to the gauge-only interpolated product (i.e., IDW) and the best performed SPP (i.e., IMERG) with increasing rain gauge densities (i. e., the number of gauges per 10^6 km^2) in the training dataset. The analyses were performed for the Chinese mainland and its five subregions. The remaining subregions including the QTP, XJ and NWC were excluded for the analyses due to the low rain gauge densities. As shown in Fig. 12, the added value of the merged product relative to the best SPP consistently exhibits an upward trend with increasing rain gauge densities, while that relative to the gauge-only interpolated product shows a downward trend. The relationships between ΔKGE and the rain gauge densities could be well represented by the logarithmic curves, with the coefficients of determination (R^2) higher than 0.70 in most of the cases. The finding agrees well with that reported by Bai et al. (2019).

We can observe from Fig. 12 that the ΔKGE values are consistently greater than zero in the Chinese Mainland and its subregions including YZ, NZ, NEC and YGP with different gauge densities in the training dataset. Meanwhile, they range from 0.09 to 0.54 (or 0.06 to 0.55) and 0.11 to 0.98 (or 0.07 to 0.97), respectively, for the merged product of RF-RF (or RF-ANN) relative to the best SPP and the gauge-only interpolated product in the subregions QTP, XJ and NWC, which are shown
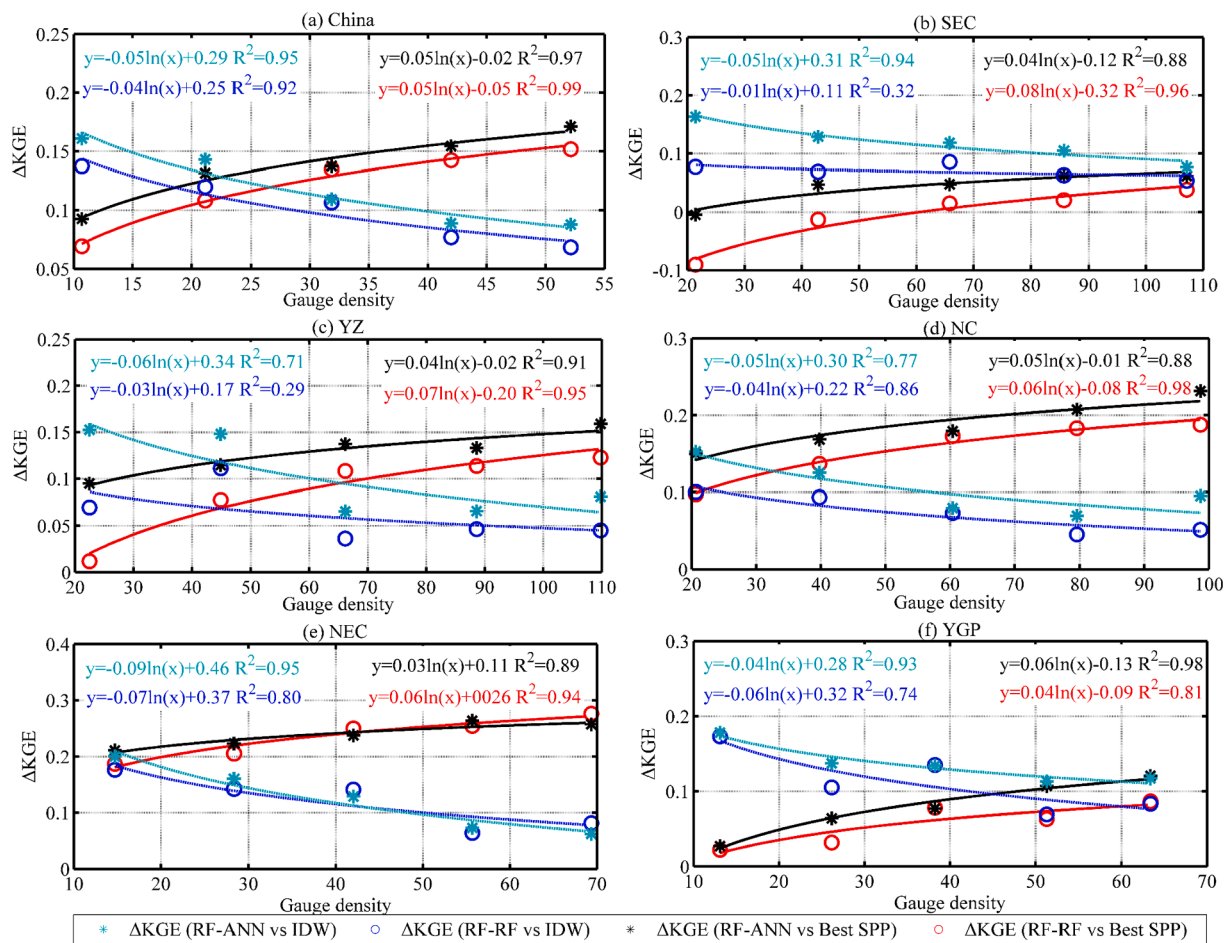
Fig. 12. Variations of the added value (i.e., ΔKGE) of the merged products of RF-ANN and RF-RF relative to the gauge-only interpolated product IDW (dark turquoise and blue colors) and IMEGR (dark black and red colors) with increasing rain gauge densities in the training dataset. The points of ΔKGE are fitted with the logarithmic curves. The six subplots show the results over the Chinese mainland and its five subregions (i.e., SEC, YZ, NC, NEC and YGP). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

here. The improvement of the precipitation estimates over the different subregions with varying rain gauge densities overall demonstrate the transferability of the proposed DML algorithm in a variety of situations. Nevertheless, we admit the new precipitation merging method is more effective in the regions with scarce or medium rain gauge densities than those with extremely low or high rain gauge densities, as reported by Wang and Lin (2015) and Bai et al. (2019). The extremely low gauge density would result in a poor representation of the training dataset, which might even induce a decrease in the performance of the merged product relative to the original SPPs, as shown in Fig. 12b. On the other hand, if the rain gauge density is high enough, the precipitation patterns could already be well captured by the gauge-based precipitation observations; and the incorporation of the precipitation information from the SPPs would have a limited ability to further improve the precipitation estimates. It should be noted that the condition "rain gauge density is extreme low or high enough" would vary with region to region, depending on the topographic and climatic features as well as the selected ML algorithm. For instance, the added value of merged product of RF-ANN relative to the gauge-only interpolated product will converge to zero in the subregion NC if the rain gauge density is higher than 403 gauges per 10^6 km^2, according to the logarithmic curves shown in Fig. 12d. However, it needs to exceed 1000 gauges per 10^6 km^2 in YGP.

As mentioned before, typically, the performance of the merged precipitation product is compared solely to the original SPPs (Yang et al., 2017; Kumar et al., 2019; Baez-Villanueva et al., 2020; Chen et al., 2020b; Wu et al., 2020). Our results, however, imply that a comparison

made to the gauge-only interpolated product is of equal importance, considering that the added value of the merged product relative to the interpolated product would decrease or even converge to zero with the increasing rain gauge densities. Moreover, we argue that more efforts are needed to investigate whether the added value of the merged product relative to the original SPPs and to the gauge-only interpolated product could be propagated to hydrological modeling (Liu et al., 2016; Ur Rahman et al., 2020).

5.3. Comparison of the machine learning algorithms

This study adopted four SML algorithms (i.e., RF, ANN, SVM and ELM) and four DML algorithms (i.e., RF-RF, RF-ANN, RF-SVM and RF-ELM) to merge multiple SPPs and gauge observations. To the best of our knowledge, this is the first attempt to compare the performances of different ML algorithms for blending satellite and gauge-based precipitation. In terms of KGE, RF performs best while SVM performs worst among the SML algorithms over the Chinese mainland. The superiority of RF to the other SML algorithms tends to be more obvious in the gauge-scarce subregions (i.e., NWC, QTP and XJ) than in the other subregions. Regarding to the DML algorithms, RF-RF and RF-ANN perform better than RF-SVM and RF-ELM. Nevertheless, there seems no consistent results at the sub-regional scale, And the performances of the SML and DML algorithms could vary with the subregions.

The DML algorithms consistently perform better than the SML algorithms in terms of KGE and CSI (precipitation threshold = 1 mm/day)

over the Chinese mainland. The median KGE for the DML-based products would increase by the ranges from 0.04 to 0.20 (or 6.09% to 41.62%) relative to the SML-based products. Meanwhile, the median CSI would increase by the ranges from 0.04 to 0.22 (or 7.05% and 52.51%). At the sub-regional level, the increases in the median KGE could be up to 6.78 or more than 10 times, and the increased in the median CSI up to 0.28 or 378.04%. The better performance of DML than SML is mainly due to the fact that the former could provide a higher classification accuracy of the 'dry/wet day' (see Table 3), which leads to a better capability in describing the temporal dynamics of precipitation, as indicated by the obviously higher Beta values (see Fig. 5d).

The increase in the median KGE could approach to 31.48% for the merged product of RF-ANN relative to the best SPP over the Chinese mainland. Baez-Villanueva-test et al. (2020) reported an improvement of their blended product of RF by about 43% in terms of KGE, in comparison to the best performed merging member. The greater improvement could be explained by the fact their rain gauges are apparently denser than ours. As shown in Fig. 12, higher rain gauge densities would lead to more added value of the merged product relative to the merging members. In the study of Wu et al. (2020), the performance improvement of their ML-based merged product is about 9% in terms of CC over the Chinese mainland, which is lower than ours that could be up to 21%. The improvement of the precipitation estimates achieved by the DML algorithms also tends to be higher than some other related studies conducted in China (Ma et al., 2018; Bai et al., 2019; Zhang et al., 2020a), which adopted the non-machine learning methods such as the linear merging approach, the geographical difference/ratio analysis, and the BMA method. The comparisons made to the previous studies prove the effectiveness of our new satellite-gauge merging method. Nevertheless, we admit that the DML approach is subjected to some deficiencies. As mentioned before, it might not be effective in the regions with extremely low or high rain gauge densities. Moreover, it might not facilitate the fusion of real-time/near-real-time data due to the fact the gauge-based observations are typically unavailable in a real-time or near-real-time manner. Additionally, the approach as well as some other ML-based methods have limited abilities to capture extreme precipitation, as pointed out by He et al. (2016) and Bhuiyan et al. (2020).

5.4. Ensemble of the DML algorithms

This study developed and adopted four DML algorithms including RF-RF, RF-ANN, RF-SVM and RF-ELM to merge multiple SPPs and gauge-observations. It is easy to raise an interesting question, i.e., whether the ensemble of the DML algorithms could further improve the precipitation estimates. In order to answer the question, we additionally produced three ensemble precipitation products based on the predictions of the four DML algorithms by using the IEVW, OORA, and OML (optimal machine leaning) methods. The IEVW and OORA methods are similar to the linear precipitation merging approaches (see Section 3.2)

with the difference that the SPPs are replaced by the ensemble members (i.e., the predictions of the DML algorithms). The OML method is to select the best performed DML algorithm to predict precipitation. The training dataset can't be simultaneously used to train and evaluate the performance of the DML algorithms, due to the issue of overfitting and each DML algorithm can provide equally good precipitation estimates in the training process. We separated the original training data into two parts: one (80%) is used to train the DML algorithms, and the other (20%) to estimate the ensemble weights of the IEVW method, and to determine the outlier and the best DML algorithm for the OORA and OML methods. As shown in Fig. 13, the ensemble precipitation products perform very comparable to the original products of the DML algorithms. This indicates that the ensemble of the DML algorithms shows little potential for the further improvement of the precipitation estimates. This might be explained by the following reasons. One is that the performance of DML is not spatially uniform and can vary significantly over different locations. The ensemble weights, the outlier and the optimal DML algorithm determined by the 20% of the training dataset might not be applicable to the gauges in the test dataset. The other is that the condensed training data (80%) will impair the performance of the DML algorithms (see Section 4.3), which will in turn affect the performance of the ensemble precipitation products.

5.5. Limitations and uncertainties

The spatial scale mismatch between the SPPs and gauge observations was neglected in the study. We assumed the point-scale and gauge-based precipitation are equal to the grid-scale and satellite-based precipitation, as done in many other studies (Ebrahimi et al., 2017; Bai et al., 2019; Kumar et al., 2019; Baez-Villanueva et al., 2020; Zhang et al., 2020a). The hypothesis, however, might not hold true, especially over the regions of complex topography, which would affect the satellite-gauge merging results, and meanwhile, bring some biases to the performance assessments of the precipitation products. In recent years, some researchers have attempted to first downscale the grid-scale SPPs to the point scale via the approaches such as the kriging and GWR (Chen et al., 2018, 2020a) or to first upscale the point-scale observation to the grid scale (Yang et al., 2017), and then carried out the satellite-gauge merging task. These operations might help to ameliorate the influences of the spatial scale mismatch, although they would produce additional uncertainties.

The training and test dataset were sampled for each subregion separately, due to the fact the rain gauges are more densely distributed in southeast China than in northwest China. The region-by-region sampling method would benefit a better representation of the training and test dataset over the Chinese Mainland, but not ensure a good representation in all of its subregions. This is because the sampling was conducted via a random manner at the subregional scale. The random sampling method, widely adopted in many previous researches (Ehsan
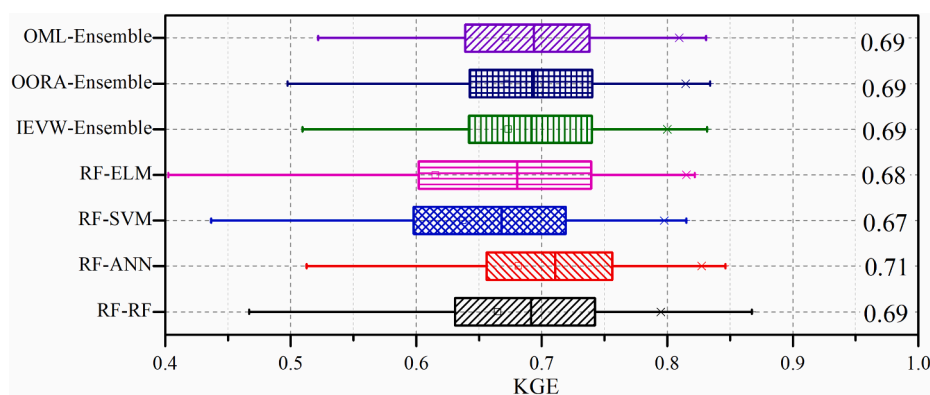

Fig. 13. Boxplots of the KGE for the ensemble precipitation products and the merged products of the DML algorithms.

Bhuiyan et al., 2019; Sharifi et al., 2019; Baez-Villanueva-test et al., 2020; Wehbe et al., 2020), might not be able to generate satisfactory samples if the rain gauges have a skewed distribution, e.g. QTP. This might bring some uncertainties to the subregional-scale results. Further efforts are needed to investigate the impacts of the sampling methods on the precipitation merging results.

This study produced a gauge-only product using the IDW interpolation algorithm to better evaluate the added value of the merged precipitation products. Although the IDW method is a popular and computationally efficient interpolation method, it has several deficiencies such as the neglect of the influence of topography on precipitation (Xu et al., 2015) and the generation of the "bull's eyes" phenomenon (Achilleos, 2008), which limits its precipitation estimation accuracy, especially in the mountainous areas with complex terrains such as the subregion QTP. Hence, the findings of the study might not be applicable if some other more advanced interpolation algorithms, e.g., MicroMet (Liston and Elder, 2006), are used for the development of the gauge-only product.

The gauge-based precipitation observations, even subjected to the quality control, might still have biases sourced from the wind-induced gauge undercatch, wetting and evaporation losses, and underestimation of trace precipitation (Ye et al., 2004; Zhang et al., 2020a). Moreover, as shown in Fig. 1, the rain gauges are unevenly distributed over the Chinse Mainland, i.e., biased toward the low-elevation and coastal areas. The uncertainties associated with rain gauges would propagate to the precipitation merging results, considering that the gauge-based observations serve as the target or response variable for the ML algorithms. Lastly, only the topographical and geographical covariates and the spatial autocorrelation information were incorporated into the precipitation merging framework. Other predictors such as the satellite-based surface soil moisture (Kumar et al., 2019) and the cloud properties (Sharifi et al., 2019), could also provide valuable information for the precipitation estimates, but they were not considered in the study. This opens for further investigations in the future.

6. Conclusions

A novel double machine learning (DML) approach was proposed to merge multiple SPPs (i.e., IMERG, SM2RAIN-ASCAT, PERISANN, and GSMap) and gauge observations over the Chinese mainland. The classification model of random forest (RF) was used together with the regression models of RF, artificial neural network (ANN), support vector machine (SVM) and extreme learning machine (ELM) to develop the DML algorithms including RF-RF, RF-ANN, RF-SVM and RF-ELM. The traditional single machine learning (SML) approach developed based solely on the regression models of RF, ANN, SVM and ELM, and three linear merging methods were also adopted in the study for the comparison purpose. We totally produced twelve precipitation products including four of the DML algorithms (i.e., RF-RF, RF-ANN, RF-SVM, and RF-ELM), four of the SML algorithms (i.e., RF, ANN, SVM, and ELM), three of the linear merging methods (i.e., IEWV, OORA, and OWA), and another one product generated via the inverse distance weighting (IDW) interpolation algorithm. The gauge-only interpolated product and the original SPPs serve as the benchmarks for the evaluations of the added value of the merged products. The gauge observation at 697 gauges were obtained over the Chinese mainland, of which 70% (i.e., 493 gauges) were used for the training of the ML algorithm or for the interpolation, while the remaining 30% (i.e., 204 gauges) for the performance evaluations. The continuous and categorical metrics including the Kling-Gupta efficiency (KGE), the probability of detection (POD), success ratio (SR), bias score (BS), and critical success index (CSI) were adopted for the performance assessments.

Results indicate that the DML algorithms achieves a better performance than the other merging methods, the gauge-only interpolation and the original SPPs over the Chinese mainland. The median KGE range 0.67–0.71 for the merged products of the DML algorithms, while they

range 0.31–0.54, 0.62, 0.54–0.55, and 0.47–0.65, respectively for the original SPPs, the gauge-only interpolated product, the linear merged products, and the SML-based products.

At the subregional scale, the DML-based products could achieve increments of the KEG by up to 0.32 (or 249.94%) and 0.22 (or 135.26%), 6.78 (or more than 10 times), respectively, in comparison to the best performed SPP, the interpolated product, and the SML-based products. The better performance of DML than SML is mainly due to that the former could achieve a higher capability in describing the temporal variations of precipitation. Regarding to the detectability of precipitation events, the DML-based products also outperform the other products with the threshold of 1 mm/day, and perform better than the original SPPs regardless of the precipitation thresholds. The increase of the median CSI (threshold = 1 mm/day) could be up to 0.19 (or 60.01%), 0.08 (or 27.56%) and 0.25 (or 378.04%), respectively, in comparison to the best performed SPP, the gauge-only interpolated product, and the SML-based products over the Chinese mainland.

Further analyses indicate that the DML-based products could perform better than the original SPPs even with a small training dataset size. Meanwhile, the added value of the merged product of DML relative to the original SPPs shows an upward trend with increasing rain gauge densities, while that relative to the gauge-only interpolated product exhibits a downward. In addition, the ensemble of the DML algorithms proves to have little potential for the further improvements of the precipitation estimates. The comparisons of the ML algorithms demonstrate that RF performs best among the four SML methods, and RF-RF and RF-ANN perform better than the other two DML algorithms (i.e., RF-SVM and RF-ELM). The major contribution of the study lies in two aspects. One is that we proposed a novel DML approach to merge multiple SPPs and gauge observations, and tested its reliability and validity over the Chinese mainland. The other is that, to our best knowledge, this is the first attempt to compare the performances of different ML algorithms in merging satellite and gauge-based precipitation, especially with the consideration of the influence of rain gauge density.

7. Codes availability

Codes of the DML and SML algorithms used to merge SPPs and gauge observations are freely available at: https://github.com/zhanglingky/MLPrecMerg. In addition to precipitation merging purposes, these programs could be easily extended to some other applications such as the downscaling of soil moisture and the hydrological predictions.

CRediT authorship contribution statement

Ling Zhang: Conceptualization, Methodology, Funding acquisition, Investigation, Software, Resources, Visualization, Writing - original draft, Writing - review & editing. **Xin Li:** Funding acquisition, Investigation, Resources, Visualization, Writing - review & editing. **Donghai Zheng:** Investigation, Formal analysis, Writing - review & editing. **Kun Zhang:** Data curation, Writing - review & editing. **Qimin Ma:** Software, Investigation. **Yanbo Zhao:** Resources, Software. **Yingchun Ge:** Validation, Writing - review & editing.

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.

like to thank the China Meteorological Data Services Center for the provision of the gauge-based precipitation observations. The authors also wish to express their gratitude to the anonymous reviewers for their valuable suggestions and comments.

References

Achilleos, G., 2008. Errors within the Inverse Distance Weighted (IDW) interpolation procedure. Geocarto Int. 23 (6), 429–449.

Baez-Villanueva-test, O.M., Zambrano-Bigiarini, M., Beck, H.E., McNamara, I., Ribbe, L., Nauditt, A., Birkel, C., Verbist, K., Giraldo-Osorio, J.D., Xuan Thinh, N., 2020. RF-MEP: A novel Random Forest method for merging gridded precipitation products and ground-based measurements. Remote Sens. Environ. 239, 111606.

Baez-Villanueva, O.M., Zambrano-Bigiarini, M., Beck, H.E., McNamara, I., Ribbe, L., Nauditt, A., Birkel, C., Verbist, K., Giraldo-Osorio, J.D., Xuan Thinh, N., 2020. RF-MEP: A novel Random Forest method for merging gridded precipitation products and ground-based measurements. Remote Sens. Environ. 239, 111606.

Bai, X., Wu, X., Wang, P., 2019. Blending long-term satellite-based precipitation data with gauge observations for drought monitoring: Considering effects of different gauge densities. J. Hydrol. 577, 124007.

Beck, H.E., Wood, E.F., McVicar, T.R., Zambrano-Bigiarini, M., Alvarez-Garreton, C., Baez-Villanueva, O.M., Sheffield, J., Karger, D.N., 2020. Bias correction of global high-resolution precipitation climatologies using streamflow observations from 9372 catchments. J. Clim. 33 (4), 1299–1315.

Beck, H.E., Wood, E.F., Pan, M., Fisher, C.K., Miralles, D.G., van Dijk, A.I.J.M., McVicar, T.R., Adler, R.F., 2019. MSWEP V2 global 3-hourly 0.1° precipitation: methodology and quantitative assessment. Bull. Am. Meteorol. Soc. 100 (3), 473–500.

Belabid, N., Zhao, F., Brocca, L., Huang, Y., Tan, Y., 2019. Near-real-time flood forecasting based on satellite precipitation products. Remote Sens. 11 (3), 252.

Bharti, V., Singh, C., 2015. Evaluation of error in TRMM 3B42V7 precipitation estimates over the Himalayan region. J. Geophys. Res.: Atmos. 120 (24), 12458–12473.

Bhuiyan, M.A.E., Nikolopoulos, E.I., Anagnostou, E.N., Quintana-Seguí, P., Barella-Ortiz, A., 2018. A nonparametric statistical technique for combining global precipitation datasets: development and hydrological evaluation over the Iberian Peninsula. Hydrol. Earth Syst. Sci. 22 (2), 1371–1389.

Bhuiyan, M.A.E., Yang, F., Biswas, N.K., Rahat, S.H., Neelam, T.J., 2020. Machine learning-based error modeling to improve GPM IMERG precipitation product over the brahmaputra river basin. Forecasting 2 (3), 248–266.

Boucher, M.A., Quilty, J., Adamowski, J., 2020. Data assimilation for streamflow forecasting using Extreme Learning Machines and Multilayer Perceptrons. Water Resour. Res. 56: e2019WR026226.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.

Breugem, A.J., Wesseling, J.G., Oostindie, K., Ritsema, C.J., 2020. Meteorological aspects of heavy precipitation in relation to floods – An overview. Earth Sci. Rev. 204, 103171.

Brocca, L., Filippucci, P., Hahn, S., Ciabatta, L., Massari, C., Camici, S., Schüller, L., Bojkov, B., Wagner, W., 2019. SM2RAIN–ASCAT (2007–2018): global daily satellite rainfall data from ASCAT soil moisture observations. Earth Syst. Sci. Data 11 (4), 1583–1601.

Brocca, L., Moramarco, T., Melone, F., Wagner, W., 2013. A new method for rainfall estimation through soil moisture observations. Geophys. Res. Lett. 40 (5), 853–858.

Brodeur, Z.P., Steinschneider, S., 2020. Spatial Bias In Medium-Range Forecasts Of Heavy Precipitation In The Sacramento River Basin: Implications For Water Management. J. Hydrometeorol. 21 (7), 1405–1423.

Camera, C., Bruggeman, A., Hadjinicolaou, P., Pashiardis, S., Lange, M.A., 2014. Evaluation of interpolation techniques for the creation of gridded daily precipitation (1× 1 km2); Cyprus, 1980–2010. J. Geophys. Res.: Atmos. 119 (2), 693–712.

Chang, C.-C., Lin, C.-J., 2011. LIBSVM: A library for support vector machines. 2(3 %J ACM Trans. Intell. Syst. Technol.): Article 27.

Chao, L., Zhang, K., Li, Z., Zhu, Y., Wang, J., Yu, Z., 2018. Geographically weighted regression based methods for merging satellite and gauge precipitation. J. Hydrol. 558, 275–289.

Chen, F., Gao, Y., 2018. Evaluation of precipitation trends from high-resolution satellite precipitation products over Mainland China. Clim. Dyn. 51 (9–10), 3311–3331.

Chen, F., Gao, Y., Wang, Y., Li, X., 2020a. A downscaling-merging method for high-resolution daily precipitation estimation. J. Hydrol. 581, 124414.

Chen, F., Li, X., 2016. Evaluation of IMERG and TRMM 3B43 monthly precipitation products over mainland China. Rem. Sens. 8 (6), 472.

Chen, S.-T., Yu, P.-S., Tang, Y.-H., 2010. Statistical downscaling of daily precipitation using support vector machines and multivariate analysis. J. Hydrol. 385 (1–4), 13–22.

Chen, S., Hong, Y., Cao, Q., Gourley, J.J., Kirstetter, P.-E., Yong, B., Tian, Y., Zhang, Z., Shen, Y., Hu, J., Hardy, J., 2013. Similarity and difference of the two successive V6 and V7 TRMM multisatellite precipitation analysis performance over China. J. Geophys. Res.: Atmos. 118 (23), 13060–13074.

Chen, S., Xiong, L., Ma, Q., Kim, J.-S., Chen, J., Xu, C.-Y., 2020b. Improving daily spatial precipitation estimates by merging gauge observation with multiple satellite-based precipitation products based on the geographically weighted ridge regression method. J. Hydrol. 589, 125156.

Chen, Y., Huang, J., Sheng, S., Mansaray, L.R., Liu, Z., Wu, H., Wang, X., 2018. A new downscaling-integration framework for high-resolution monthly precipitation estimates: Combining rain gauge observations, satellite-derived precipitation data and geographical ancillary data. Rem. Sens. Environ. 214, 154–172.

Du, Y., Xie, Z.Q., Miao, Q., 2020. Spatial Scales of Heavy Meiyu Precipitation Events in Eastern China and Associated Atmospheric Processes. Geophys. Res. Lett. 46, e2020GL087086.

Duan, Z., Bastiaanssen, W.G.M., 2013. First results from Version 7 TRMM 3B43 precipitation product in combination with a new downscaling–calibration procedure. Rem. Sens. Environ. 131, 1–13.

Ebrahimi, S., Chen, C., Chen, Q., Zhang, Y., Ma, N., Zaman, Q., 2017. Effects of temporal scales and space mismatches on the TRMM 3B42 v7 precipitation product in a remote mountainous area. Hydrol. Process. 31 (24), 4315–4327.

Ehsan Bhuiyan, M.A., Nikolopoulos, E.I., Anagnostou, E.N., 2019. Machine learning-based blending of satellite and reanalysis precipitation datasets: a multiregional tropical complex terrain evaluation. J. Hydrometeorol. 20 (11), 2147–2161.

Gardner, M.W., Dorling, S.R., 1998. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. Atmos. Environ. 32 (14), 2627–2636.

Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. J. Hydrol. 377 (1–2), 80–91.

Gupta, V., Jain, M.K., Singh, P.K., Singh, V., 2019. An assessment of global satellite-based precipitation datasets in capturing precipitation extremes: A comparison with observed precipitation dataset in India. Int. J. Climatol. 40, 3667–3688.

He, X., Chaney, N.W., Schleiss, M., Sheffield, J., 2016. Spatial downscaling of precipitation using adaptable random forests. Water Resour. Res. 52 (10), 8217–8237.

Hengl, T., Nussbaum, M., Wright, M.N., Heuvelink, G.B.M., Gräler, B., 2018. Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. PeerJ 6, e5518.

Huang, G., Zhou, H., Ding, X., Zhang, R., 2012. Extreme learning machine for regression and multiclass classification. IEEE Trans. Syst. Man Cybernet. Part B (Cybernet.) 42 (2), 513–529.

Huffman, G.J., Bolvin, D., Braithwaite, D., Hsu, K., Joyce, R., Xie, P., 2014. Integrated Multi-satellitE Retrievals for GPM (IMERG), version 4.4. NASA's Precipitation Processing Center, accessed 31 March, 2015, ftp://arthurhou.pps.eosdis.nasa.gov/gpmdata/.

Huffman, G.J., Bolvin, D.T., Nelkin, E.J., Tan, J., 2019. Integrated Multi-satellitE Retrievals for GPM (IMERG) Technical Documentation, https://docserver.gesdisc.eosdis.nasa.gov/public/project/GPM/IMERG_doc.06.pdf.

Kling, H., Fuchs, M., Paulin, M., 2012. Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. J. Hydrol. 424–425, 264–277.

Kubota, T., Shige, S., Hashizume, H., Aonashi, K., Takahashi, N., Seto, S., Hirose, M., Takayabu, Y.N., Ushio, T., Nakagawa, K., Iwanami, K., Kachi, M., Okamoto, K.i., 2007. Global precipitation map using satellite-borne microwave radiometers by the GSMaP project: production and validation. IEEE Trans. Geosci. Remote Sens. 45(7), 2259–2275.

Kumar, A., Ramsankaran, R., Brocca, L., Munoz-Arriola, F., 2019. A machine learning approach for improving near-real-time satellite-based rainfall estimates by integrating soil moisture. Remote Sens. 11 (19), 2221.

Li, M., Shao, Q., 2010. An improved statistical approach to merge satellite rainfall estimates and raingauge data. J. Hydrol. 385 (1–4), 51–64.

Li, T., Shen, H., Yuan, Q., Zhang, X., Zhang, L., 2017. Estimating ground-level PM2.5 by fusing satellite and station observations: a geo-intelligent deep learning approach. Geophys. Res. Lett. 44 (23), 11985–11993.

Lin, P., Pan, M., Allen, G.H., Frasson, R.P., Zeng, Z., Yamazaki, D., Wood, E.F., 2020. Global estimates of reach-level bankfull river width leveraging big data geospatial analysis. Geophys. Res. Lett. 47 (7).

Liston, G.E., Elder, K., 2006. A meteorological distribution system for high-resolution terrestrial modeling (MicroMet). J. Hydrometeorol. 7 (2), 217–234.

Liu, S., Xu, Z., Song, L., Zhao, Q., Ge, Y., Xu, T., Ma, Y., Zhu, Z., Jia, Z., Zhang, F., 2016. Upscaling evapotranspiration measurements from multi-site to the satellite pixel scale over heterogeneous land surfaces. Agric. For. Meteorol. 230–231, 97–113.

Ma, Y., Hong, Y., Chen, Y., Yang, Y., Tang, G., Yao, Y., Long, D., Li, C., Han, Z., Liu, R., 2018. Performance of optimally merged multisatellite precipitation products using the dynamic bayesian model averaging scheme over the Tibetan Plateau. J. Geophys. Res.: Atmos. 123 (2), 814–834.

Manz, B., Buytaert, W., Zulkafli, Z., Lavado, W., Willems, B., Robles, L.A., Rodríguez-Sánchez, J.-P., 2016. High-resolution satellite-gauge merged precipitation climatologies of the Tropical Andes. J. Geophys. Res.: Atmos. 121 (3), 1190–1207.

Markonis, Y., Papalexiou, S.M., Martinkova, M., Hanel, M., 2019. Assessment of water cycle intensification over land using a multi-source global gridded precipitation dataset. J. Geophys. Res.: Atmos. 124, 11175–11187.

Mastrantonas, N., Bhattacharya, B., Shibuo, Y., Rasmy, M., Espinoza-Dávalos, G., Solomatine, D., 2019. Evaluating the benefits of merging near-real-time satellite precipitation products: a case study in the Kinu Basin Region, Japan. J. Hydrometeorol. 20 (6), 1213–1233.

Merz, R., Tarasova, L., Basso, S., 2020. Parameter's Controls of Distributed Catchment Models—How Much Information is in Conventional Catchment Descriptors? Water Resour. Res. 56(2), e2019WR026008.

Munier, S., Aires, F., Schlaffer, S., Prigent, C., Papa, F., Maisongrande, P., Pan, M., 2014. Combining data sets of satellite-retrieved products for basin-scale water balance study: 2. Evaluation on the Mississippi Basin and closure correction model. J. Geophys. Res.: Atmos. 119 (21), 12100–12116.

Nguyen, P., Shearer, E.J., Tran, H., Ombadi, M., Hayatbini, N., Palacios, T., Huynh, P., Braithwaite, D., Updegraff, G., Hsu, K., Kuligowski, B., Logan, W.S., Sorooshian, S., 2019. The CHRS Data Portal, an easily accessible public repository for PERSIANN global satellite precipitation data. Sci. Data 6 (1), 180296.

Nie, S., Luo, Y., Wu, T., Shi, X., Wang, Z., 2015. A merging scheme for constructing daily precipitation analyses based on objective bias-correction and error estimation techniques. J. Geophys. Res.: Atmos. 120 (17), 8671–8692.

Pellet, V., Aires, F., Munier, S., Fernández Prieto, D., Jordá, G., Dorigo, W.A., Polcher, J., Brocca, L., 2019. Integrating multiple satellite observations into a coherent dataset to monitor the full water cycle – application to the Mediterranean region. Hydrol. Earth Syst. Sci. 23 (1), 465–491.

Rahman, K.U., Shang, S., Shahid, M., Wen, Y., Khan, Z., 2020. Application of a dynamic clustered bayesian model averaging (DCBA) algorithm for merging multisatellite precipitation products over Pakistan. J. Hydrometeorol. 21 (1), 17–37.

Roebber, P.J., 2009. Visualizing multiple measures of forecast quality. Weather Forecasting 24 (2), 601–608.

Schmidt, L., Heße, F., Attinger, S., Kumar, R., 2020. Challenges in applying machine learning models for hydrological inference: A case study for flooding events across Germany. Water Resour. Res., 56, e2019WR025924.

Sharifi, E., Saghafian, B., Steinacker, R., 2019. Downscaling satellite precipitation estimates with multiple linear regression, artificial neural networks, and spline interpolation techniques. J. Geophys. Res.: Atmos. 124 (2), 789–805.

Shen, Y., Xiong, A., Hong, Y., Yu, J., Pan, Y., Chen, Z., Saharia, M., 2014. Uncertainty analysis of five satellite-based precipitation products and evaluation of three optimally merged multi-algorithm products over the Tibetan Plateau. Int. J. Remote Sens. 35 (19), 6843.

Shen, Y., Xiong, A., Wang, Y., Xie, P., 2010. Performance of high-resolution satellite precipitation products over China. J. Geophys. Res. 115, D02114.

Sorooshian, S., Hsu, K.-L., Gao, X., Gupta, H.V., Imam, B., Braithwaite, D., 2000. Evaluation of PERSIANN system satellite-based estimates of tropical rainfall. Bull. Am. Meteorol. Soc. 81 (9), 2035–2046.

Tesfa, T.K., Leung, L.R., Ghan, S.J., 2020. Exploring topography-based methods for downscaling subgrid precipitation for use in earth system models. J. Geophys. Res.: Atmos. 125 (5).

Thornton, P.E., Running, S.W., White, M.A., 1997. Generating surfaces of daily meteorological variables over large regions of complex terrain. J. Hydrol. 190 (3), 214.

Ur Rahman, K., Shang, S., Shahid, M., Wen, Y., 2020. Hydrological evaluation of merged satellite precipitation datasets for streamflow simulation using SWAT: A case study of Potohar Plateau, Pakistan. J. Hydrol. 587, 125040.

Vapnik, V.N., 1998. Statistical Learning Theory. John Wiley, New Yor.

Wang, L., Li, X., Ma, C., Bai, Y., 2019. Improving the prediction accuracy of monthly streamflow using a data-driven model based on a double-processing strategy. J. Hydrol. 573, 733–745.

Wang, X.L., Lin, A., 2015. An algorithm for integrating satellite precipitation estimates with in situ precipitation data on a pentad time scale. J. Geophys. Res.: Atmos. 120 (9), 3728–3744.

Wehbe, Y., Temimi, M., Adler, R.F., 2020. Enhancing precipitation estimates through the fusion of weather radar, satellite retrievals, and surface parameters. Rem. Sens. 12 (8), 1342.

Wu, H., Yang, Q., Liu, J., Wang, G., 2020. A spatiotemporal deep fusion model for merging satellite and gauge precipitation in China. J. Hydrol. 584, 124664.

Xie, P., Xiong, A.-Y., 2011. A conceptual model for constructing high-resolution gauge-satellite merged precipitation analyses. J. Geophys. Res.: Atmos. 116 (D21).

Xu, L., Chen, N., Moradkhani, H., Zhang, X., Hu, C., 2020. Improving global monthly and daily precipitation estimation by fusing gauge observations, remote sensing and reanalysis datasets. Water Resour. Res.

Xu, R., Tian, F., Yang, L., Hu, H., Lu, H., Hou, A., 2017. Ground validation of GPM IMERG and TRMM 3B42V7 rainfall products over southern Tibetan Plateau based on a high-density rain gauge network. J. Geophys. Res.: Atmos. 122 (2), 910–924.

Xu, W., Zou, Y., Zhang, G., Linderman, M., 2015. A comparison among spatial interpolation techniques for daily rainfall data in Sichuan Province, China. Int. J. Climatol. 35 (10), 2898–2907.

Yang, Y., Luo, Y., 2014. Using the back propagation neural network approach to bias correct TMPA data in the arid region of northwest China. J. Hydrometeorol. 15 (1), 459–473.

Yang, Z., Hsu, K., Sorooshian, S., Xu, X., Braithwaite, D., Zhang, Y., Verbist, K.M.J., 2017. Merging high-resolution satellite-based precipitation fields and point-scale rain gauge measurements-A case study in Chile. J. Geophys. Res.: Atmos. 122 (10), 5267–5284.

Ye, B., Yang, D., Ding, Y., Han, T., Koike, T., 2004. A bias-corrected precipitation climatology for China. J. Hydrometeorol. 5 (6), 1147–1160.

Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., Xu, H., Tan, W., Yang, Q., Wang, J., Gao, J., Zhang, L., 2020. Deep learning in environmental remote sensing: Achievements and challenges. Rem. Sens. Environ. 241, 111716.

Zhang, L., Li, X., Cao, Y., Nan, Z., Wang, W., Ge, Y., Wang, P., Yu, W., 2020a. Evaluation and integration of the top-down and bottom-up satellite precipitation products over mainland China. J. Hydrol. 581, 124456.

Zhang, L., Ren, D., Nan, Z., Wang, W., Zhao, Y., Zhao, Y., Ma, Q., Wu, X., 2020b. Interpolated or satellite-based precipitation? Implications for hydrological modeling in a meso-scale mountainous watershed on the Qinghai-Tibet Plateau. J. Hydrol. 583, 124629.

Zhang, X., Tang, Q., 2015. Combining satellite precipitation and long-term ground observations for hydrological monitoring in China. J. Geophys. Res.: Atmos. 120 (13), 6426–6443.

Zhao, T., Yatagai, A., 2014. Evaluation of TRMM 3B42 product using a new gauge-based analysis of daily precipitation over China. Int. J. Climatol. 34 (8), 2749–2762.

Zhong, R., Chen, X., Lai, C., Wang, Z., Lian, Y., Yu, H., Wu, X., 2019. Drought monitoring utility of satellite-based precipitation products across mainland China. J. Hydrol. 568, 343–359.