

Article

# Towards an Accurate and Reliable Downscaling Scheme for High-Spatial-Resolution Precipitation Data

Honglin Zhu <sup>1</sup>, Huizeng Liu <sup>2</sup> , Qiming Zhou <sup>1,\*</sup>  and Aihong Cui <sup>1</sup><sup>1</sup> Department of Geography, Hong Kong Baptist University, Hong Kong, China<sup>2</sup> Institute for Advanced Study, Shenzhen University, Shenzhen 518060, China

\* Correspondence: qiming@hkbu.edu.hk

**Abstract:** Accurate high-spatial-resolution precipitation is significantly important in hydrological and meteorological modelling, especially in rain-gauge-sparse areas. Some methods and strategies have been applied for satellite-based precipitation downscaling, residual correction and precipitation calibration. However, which downscaling scheme can provide reliable high-resolution precipitation efficiently remains unanswered. To address this issue, this study aimed to present a framework combining the machine learning downscaling algorithm and post-process procedures. Firstly, four ML-based models, namely support vector regression, random forest, spatial random forest (SRF) and eXtreme gradient boosting (XGBoost), were tested for downscaling and compared with conventional downscaling methods. Then, the effectiveness of the residual correction process using ordinary Kriging and the calibration process using the geographical difference analysis (GDA) method was investigated. The results showed that the ML-based methods had better performance than the conventional regression and interpolation approaches. The SRF and XGBoost outperformed others in generating accurate precipitation estimation with a high resolution. The GDA calibration process significantly improved the downscaled results. However, the residual correction process decreased the downscaling performance of the ML-based models. Combining the SRF or XGBoost downscaling algorithm with the GDA calibration method could be a promising downscaling scheme for precipitation data. The scheme could be used to generate high-resolution precipitation, especially in areas urgently requiring data, which would benefit regional water resource management and hydrological disaster prevention.



**Citation:** Zhu, H.; Liu, H.; Zhou, Q.; Cui, A. Towards an Accurate and Reliable Downscaling Scheme for High-Spatial-Resolution Precipitation Data. *Remote Sens.* **2023**, *15*, 2640. <https://doi.org/10.3390/rs15102640>

Academic Editor: Yuanjian Yang

Received: 16 March 2023

Revised: 27 April 2023

Accepted: 16 May 2023

Published: 18 May 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

As the primary force of the hydrological cycle and energy balance, precipitation is a significant variable in meteorological and hydrological modelling [1–4]. The spatial and temporal distribution of precipitation are two decisive factors in global water cycle and climate change studies [5–8]. However, precipitation is one of the most difficult meteorological components to estimate due to its high spatiotemporal variations [9]. Although in situ rainfall gauges can provide reliable observations, the limited and uneven distribution of rain gauge stations makes it difficult to reflect the spatial pattern of precipitation. Therefore, acquiring accurate and high-resolution precipitation data still remains challenging, especially in rain-gauge-sparse areas.

Alternatively, remote sensing could also be used to estimate precipitation with wide spatial coverage. Many satellite precipitation estimates (SPEs) have been produced, such as the Tropical Rainfall Measuring Mission [10,11], Climate Hazards Group Infrared Precipitation with Station data [8], the Integrated Multi-Satellite Retrievals for Global Precipitation Measurement mission [12] and the Global Precipitation Climatology Project [13,14]. One of the satellite precipitation products is the Precipitation Estimation from Remotely

Sensed Information using Artificial Neural Networks-Climate Data Record (PERSIANN-CDR) [15,16], which is widely used in many examples of hydrological modelling and has been reported to have good performance in mainland China [17]. However, its resolutions, varying from  $0.25^{\circ}$  to  $0.10^{\circ}$ , are too coarse to be applied in regional-scale studies. Therefore, spatial downscaling is required to produce high-resolution precipitation to make it more applicable on a local scale.

There are mainly two downscaling techniques, dynamic and statistical downscaling, that have been widely used, as they rely on less computational costs and no physical assumptions [18]. They are implemented by first developing regression models between precipitation and environmental factors with a coarse resolution, and then applying the built models to finer environmental variables to produce high-resolution precipitation. Many regression models have been applied to drive the hydrometeorological variables as predictors for providing downscaling results [19]. For example, the relationship between the precipitation and normalised difference vegetation index (NDVI) at different spatial resolutions was first explored using the exponential model (ER) [20]. Jia et al. [21] used the multiple linear regression (MLR) model to develop the correlation between precipitation, vegetation and topography. Xu et al. [22] proposed a new geographically weighted regression (GWR) model to estimate high-resolution precipitation based on the NDVI and DEM.

In addition to the simple statistical regression algorithms, some advanced ML algorithms have also been adopted for downscaling, and have been found to outperform other conventional approaches [23–25]. The support vector machine (SVM) was first used by Chen et al., (2010) [26] to spatially downscale the general circulation models (GCMs) of precipitation. He et al. (2016) [27] proposed an adoptable RF-based approach for precipitation downscaling, which employed two independent RFs and yielded better estimation for extreme precipitation. Jing et al., (2018) [2] applied the classification and regression trees (CARTs) algorithm in the downscaling of TRMM precipitation and reported its better performance than the k-nearest neighbours (KNNs) method. Machine learning algorithms have been utilised not only as stand-alone tools, but also in conjunction with other downscaling methods. For example, Devak et al., (2015) [28] developed a dynamic downscaling framework by integrating the KNN and SVM techniques. Yan et al. [29] presented a downscaling and merging scheme based on the RF and Cokriging model, and achieved better accuracy than the original global precipitation measurement mission (GPM) precipitation. Pour et al., (2016) [30] developed a hybrid approach using the RF in classification and SVM in regression to downscale the daily rainfall. The exploration of machine learning algorithms in precipitation downscaling research indicated their superior performance.

Following the diverse application of machine learning algorithms to downscale precipitation, some studies have evaluated and compared their performances in downscaling. Raje and Mujumdar (2011) [31] compared the performance of three downscaling methods, namely, conditional random forest (CRF), KNN, and SVM for downscaling precipitation from the Canadian global climate model, in which CRF and KNN models performed slightly better than the SVM model in reproducing high-resolution precipitation. Nasseri et al., (2013) [32] implemented the assessment of four downscaling methods, including the cubic-order multivariate adaptive regression splines (MARS), model tree (MT), KNN and genetic-algorithm-optimised SVM (GA-SVM), and the combination of the MT and MARS methods produced more accurate results. Sharifi et al., (2018) [33] evaluated three downscaling techniques (MLR, artificial neural networks (ANNs) and spline interpolation methods) based on the relationships between GPM (IMERG) and cloud properties in northeast Austria. Ghorbanpour et al., (2021) [34] compared the performance of SVM, RF, GWR, MLR and exponential regression on downscaling TRMM precipitation, and the SVM algorithm demonstrated the highest performance.

The existing comparative studies have demonstrated assessments mainly focusing on the RF and SVM algorithms. However, other advanced algorithms, such as spatial random forest (SRF) and eXtreme gradient boosting (XGBoost), have been gradually developed

and used in downscaling. Chen et al. [9] incorporated spatial autocorrelation into the RF algorithm and proposed the SRF for precipitation downscaling. XGBoost has not been reported in the downscaling of precipitation data, but it has been successfully applied in the downscaling of terrestrial water storage from the Gravity Recovery and Climate Experiment (GRACE) satellite, indicating its great potential in precipitation downscaling [35,36]. RF and SVM are the representatives of the bagging method and kernel methods, respectively, and have been widely used in the spatial downscaling of satellite precipitation estimates. Nevertheless, the application of the spatial extension of RF (SRF) and boosting methods such as XGBoost were not investigated in this field. Therefore, the effectiveness of these ML approaches to downscale the satellite precipitation data will be evaluated in this study for the best-performing downscaling algorithm.

Another key question is whether residual correction and calibration are essential to improve the first results of downscaling. The downscaling result contained the residuals that were not captured by the downscaling models, and required further residual correction and a calibration process. Residual correction considers the difference between the simulated values with a coarse resolution and the original satellite data [37]. Some studies have indicated that residual correction could improve the downscaled results [16,21,22,24]. For example, residual correction greatly corrected the precipitation bias obtained from the spline interpolation method [24], classification and regression tree [16,21] and the k-nearest neighbours [22], and thus it was considered as an essential step for downscaling. However, there were also some studies that came to the opposite conclusion. Duan and Bastiaanssen [38] suggested that the employment of residual correction reduced the performance of the multiple linear regression model for downscaling. Xu et al. [22] also found that residual correction reduced the estimate performance of the GWR model. Thus, the significance of residual correction in improving the downscaled results from ML models will be investigated.

The calibration process merges in situ observations with downscaled precipitation [39]. Geographical difference analysis (GDA) [38,39] has been proposed for improving the downscaled results. GDA calibration computed the difference between downscaled precipitation and the in situ observations at each rain gauge, and the difference was interpolated into high spatial resolution and added to the uncalibrated results. GDA can minimise the difference between the downscaled against the observed data, and it was more effective than other calibration methods, such as regression analysis [34,39], especially in areas with scanty distribution. Therefore, this study will explore whether or not the combination of the GDA and ML-based models is effective in downscaling precipitation data.

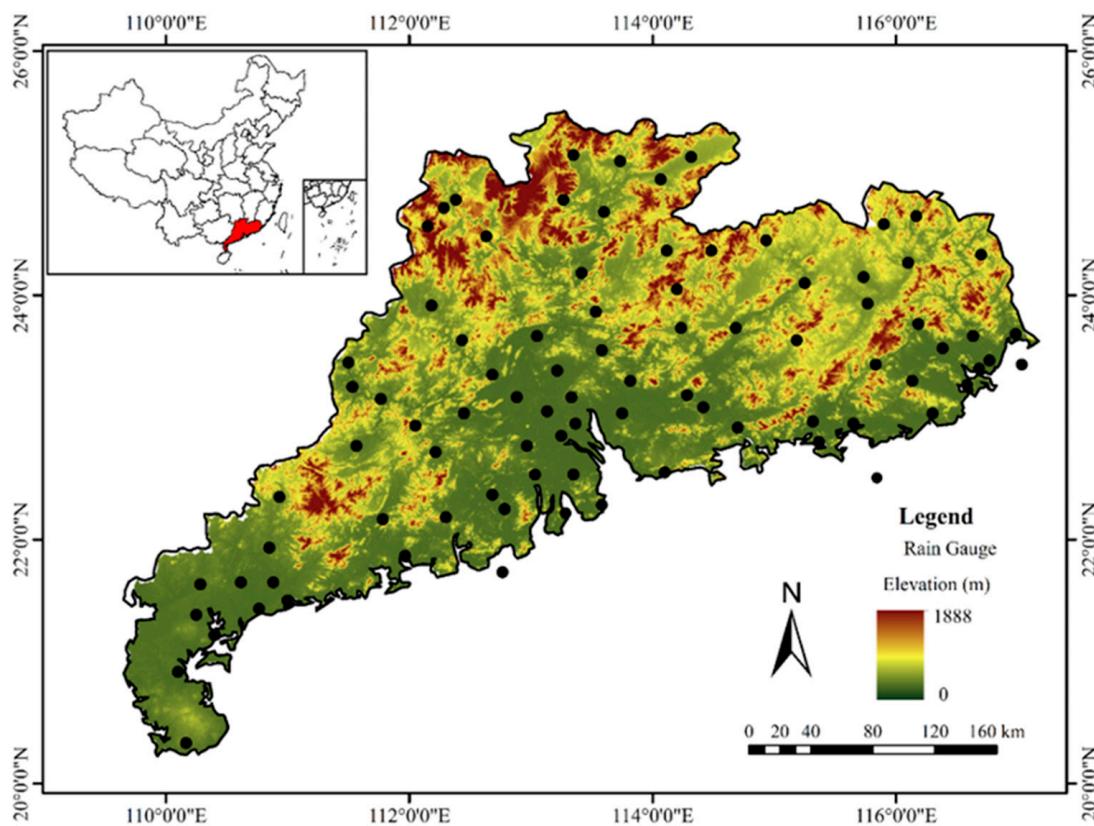
To answer these aforementioned research questions, this study aims to identify an effective downscaling scheme. The downscaling scheme was determined by the following: (1) systematically evaluating the ML-based spatial downscaling methods and (2) investigating the contribution of the residual correction and calibration process on the downscaled results. The results from this study can address the significance and contribution of ML-based downscaling models, calibration and the residual correction process, and the downscaling scheme is promising in providing accurate and reliable precipitation data with a high resolution.

## 2. Materials and Methods

### 2.1. Study Area

This study was undertaken in Guangdong Province in China, covering a total area of 179,800 km<sup>2</sup> between 109° and 118°E and 20° and 26°N. As shown in Figure 1, the study area is dominated by mountainous and hilly terrain, ranging from high in the north (1888 m above sea level) to low in the south (0 m). About 33% of the whole district is mountains, and hills and plains account for 25% and 22%, respectively [40]. Dominated by the East Asian Monsoon, the study area crosses three climatic sub-zones, i.e., the middle subtropics, the south subtropics and the tropics. The climate is characterised by a warm and relatively dry winter and hot and wet summer, with average annual precipitation ranging from 1366

to 2343 mm [41]. The abundance and high spatial instability of the rainfall make this region an appropriate choice for the assessment of precipitation downscaling and estimation.



**Figure 1.** Topography, rain gauges and geographic location of Guangdong Province in China.

## 2.2. Dataset and Pre-Processing

Table 1 shows the rain gauge observation, satellite dataset and other environmental predictors used in this study. The PERSIANN-CDR is a real-time satellite product and provides the precipitation data at  $0.25^{\circ}$  spatial resolution from 1983 for the space coverage of  $60^{\circ}\text{S}$ – $60^{\circ}\text{N}$  [16]. For this study, the annual precipitation data of PERSIANN-CDR from 2006 to 2010 were used. The observed precipitation of 86 meteorological stations (Figure 1) was provided by the National Meteorological Information Center with data quality control [42]. Daily records from January 2006 to December 2010 were used and annual precipitation at each site was calculated by the sum of the daily records. In the rare cases of missing data at some stations, observations from the 10 nearest surrounding stations were averaged as the recordings.

The DEM data were obtained from the Shuttle Radar Topography Mission (SRTM) by NASA and the National Mapping Agency [43,44]. The SRTM DEM with a spatial resolution of 90 m was resampled to  $0.01^{\circ}$  by averaging the values in the pixels. The DEM derivatives involving slope and aspect were obtained based on the SRTM DEM. In our study, NDVI was used, and the dataset was supplied by the NDVI3g by the GIMMS working group using the Advanced Very-High-Resolution Radiometer (AVHRR) data [45–47]. The data were processed using the maximum value composite (MVC) method and synthesised every fortnight (24 times per year), with a spatial resolution of  $1/12^{\circ}$  (8 km). The annual NDVI was calculated by averaging the monthly NDVI value. The MODIS 8-day Land Surface Temperature (LST) product (MOD11A2) with a spatial resolution of 1 km was provided from NASA Earth Data [48]. The annual LST values were obtained via temporal averaging.

**Table 1.** Datasets used in this study.

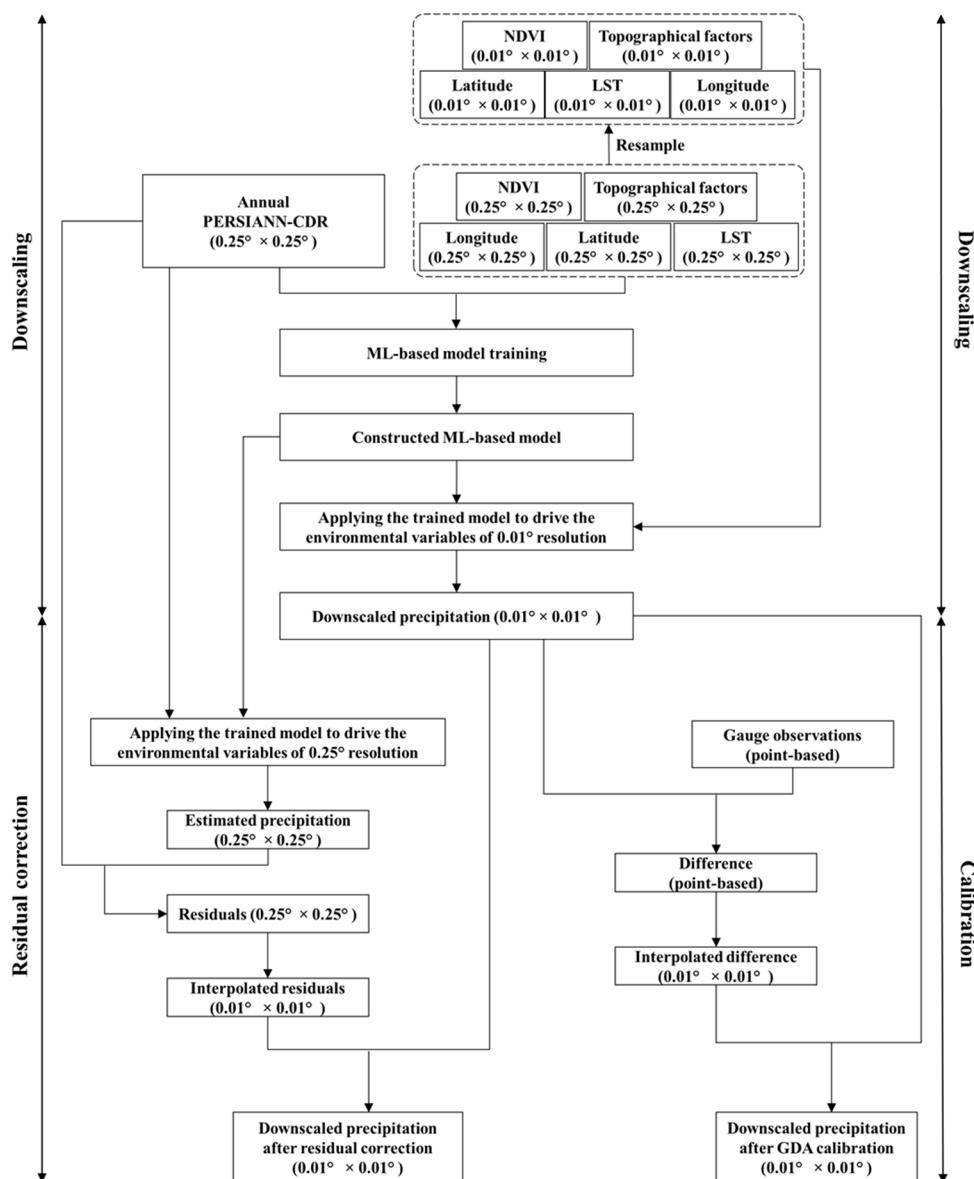
Data Type	Dataset	Spatial Resolution	Temporal Resolution	Source
Meteorological data	PERSIANN-CDR rainfall gauge observation	25 km Point	Annual Daily	<a href="http://chrs.web.uci.edu/persiann">http://chrs.web.uci.edu/persiann</a> (accessed on 1 December 2021) <a href="http://data.cma.cn/data/detail/dataCode/">http://data.cma.cn/data/detail/dataCode/</a> (accessed on 1 December 2021)
	SRTM DEM slope, aspect LST	90 m 90 m 1 km	- - 8 d	<a href="https://doi.org/10.5066/F7PR7TFT">https://doi.org/10.5066/F7PR7TFT</a> (accessed on 1 December 2021) Derived from SRTM DEM <a href="https://doi.org/10.5067/MODIS/MOD11A2.006">https://doi.org/10.5067/MODIS/MOD11A2.006</a> (accessed on 1 December 2021)
Land surface data	GIMMS NDVI3g	8 km	15 d	The National Center for Atmospheric Research

### 2.3. Downscaling of PERSIANN-CDR Precipitation

The statistical downscaling techniques were built on the assumption that the correlation between PERSIANN-CDR precipitation and environmental variables conducted with coarse resolution ( $0.25^\circ \times 0.25^\circ$ ) could be equally applied at a fine resolution ( $1 \text{ km} \times 1 \text{ km}$ ) [38,49–51]. In this study, the models considered the NDVI, DEM and LST as predictors, which was suggested by [9,33,34]. The relationship between NDVI and precipitation has been widely introduced to downscaling models because vegetation types have influenced the humidity and moist convention strongly [21,38,52]. Precipitation, as a prediction, was largely influenced by topographical factors such as elevation, slope and aspect [22,51]. Topographical factors have been adopted to estimate precipitation. LST both in daytime and at night were important responses to precipitation [53,54]. Thus, LST was used as a contributing factor to improve the accuracy of the downscaling process. In addition, the geographic location information, i.e., longitude and latitude, were also considered as input factors to drive the downscaling models [2,55].

Four ML-based algorithms (SVR, RF, SRF and XGBoost) were considered as candidates to develop the downscaling model. The RF and SVR algorithms were considered as classical ML models and have been used in many downscaling studies [25,56]. SRF was proposed as the extension of the RF method to deal with spatial prediction issues and had great potential in precipitation downscaling. SRF introduced the Moran's index of spatial autocorrelation as an explanatory variable into the RF model, and minimised the spatial autocorrelation of simulation residuals [57,58]. The XGBoost algorithm is an effective modified version of the gradient boosting decision tree model [59]. All of these four models have shown reasonable estimation results in the literature. In addition to the four ML-based algorithms, one interpolation approach (ordinary Kriging) and two regression models (MLR and GWR) were also employed as comparative studies for downscaling. Three variables were used in the MLR model. The GWR is a local regression method that introduced geographical location information into the regression model [22].

The procedure of ML-based downscaling methods is shown in Figure 2. It was achieved by firstly developing the regression models between precipitation and environmental variables on a coarse scale and then applying the models to the high-resolution environmental factors to generate the precipitation data with high spatial resolution. The data from 2006 to 2010 were used as the training and validation dataset, in which 80% was used to adjust the parameters of the regression model and the other 20% was applied for validation. To avoid the overfitting of the model, the five-fold cross-validation [57] method was implemented. In this study, the R package "e1071" was used for constructing the SVR model [60]. The R package "randomForest" was used for the RF model [61]. The SRF model and XGBoost model were implemented in the R package "spatialRF" [62] and "xgboost" [63], respectively. The GWR model was played in GWR 4.0 software [64].



**Figure 2.** The flowchart of the ML-based downscaling, residual correction and calibration process.

#### 2.4. Residual Correction and Calibration Framework

As shown in Figure 2, the residual correction was conducted using three steps [33,37]. First, the estimated precipitation with  $0.25^\circ$  resolution using the constructed models was obtained and subtracted from the original PERSIANN-CDR dataset to obtain the residuals with  $0.25^\circ$  resolution. The residuals with coarse resolution were then interpolated to  $0.01^\circ$  using the ordinary Kriging method. Finally, the downscaled precipitation with a high resolution was corrected by adding the residual at  $0.01^\circ$  resolution. Different from the residual correction, the calibration process considered the difference between the downscaled rainfall and observation measurements. The GDA calibration approach proposed by [39] was adopted, which was more effective than the geographical ratio analysis (GRA) [38]. The process of the GDA calibration method was demonstrated in Figure 2. Firstly, the downscaled precipitation map was extracted to the point-based data according to the location of rain gauges, which were then subtracted from the observed recordings to obtain the difference. The point-based difference was then interpolated using the inverse distance weighting (IDW) technique, and the interpolated differences were added to the downscaled precipitation as the calibrated results. The mean yearly precipitation for all 86 rain gauges was ranked from low to high, with the sequence of 1–81.

A total of 50% of the rain gauge stations with odd numbers were selected for calibration, and the remaining ones were used for validation.

### 2.5. Performance Evaluation

To evaluate the downscaling results, four indicators of correlation coefficient (CC), root mean square error (RMSE), mean absolute error (MAE) and Kling-Gupta efficiency (KGE) were used. The correlation coefficient is the ratio between the covariance of two variables and the product of their standard deviations; RMSE is the arithmetic square root of the difference between the measured precipitation value and the predicted precipitation value at the site and MAE is the average of the absolute value of the difference between the measured precipitation value and the predicted precipitation value at the site. In addition, the KGE was applied to evaluate the overall performance of the downscaled precipitation data. KGE combines three components of model errors (i.e., correlation, bias and coefficients of variation) in a balanced way [65]. The equations of the index applied are presented in Table 2.

**Table 2.** The evaluation index and the equation applied.

Evaluation Index	Equation
Correlation coefficient (CC)	$CC = \frac{\sum_{i=1}^n (E_i - \bar{E})(O_i - \bar{O})}{\sqrt{\sum_{i=1}^n (E_i - \bar{E})^2} \times \sqrt{\sum_{i=1}^n (O_i - \bar{O})^2}}$
Root mean square error (RMSE)	$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (E_i - O_i)^2}$
Mean absolute error (MAE)	$MAE = \frac{\sum_{i=1}^n  E_i - O_i }{n}$
Kling-Gupta efficiency (KGE)	$KGE = 1 - \sqrt{(cc - 1)^2 + \left(\frac{\bar{E}}{\bar{O}} - 1\right)^2 + \left(\frac{CV_E}{CV_O} - 1\right)^2}$

where  $E_i$  = the estimated precipitation at station  $i$ ,  $O_i$  = the observed precipitation at station  $I$  and  $n$  is the number of rain gauge stations.  $\bar{E}$  is the mean of the estimated precipitation.  $\bar{O}$  is the mean of the observed precipitation.  $CV_E$  is the coefficient of variation of the estimated precipitation, and  $CV_O$  is the coefficient of variation of the observed precipitation.

## 3. Results

### 3.1. Accuracy Analysis of ML-Based and Conventional Downscaled Methods

Table 3 shows the CC, MAE, RMSE and KGE of each method including the ML-based and the conventional methods based on the 86 individualistic rain gauges used for each validation year from 2006 to 2010. The accuracy of the original PERSIANN-CDR was not very high, having the lowest CC and KGE and the highest MAE and RMSE. The downscaling methods applied improved the spatial resolution as well as the accuracy. The validation results of each year differed and were related to the performance of the original satellite precipitation, indicating that the accuracy of the downscaled results was not only affected by the accuracy of the regression model, but also the quality of the original satellite precipitation data. The four ML-based methods produced better performance than the other three conventional methods in terms of KGE, indicating their better capability in fitting non-linear relationships between satellite precipitation and environmental variables.

The performance of Kriging and MLR was unsatisfactory with low CC and KGE, as the relationship between precipitation and environmental factors cannot be properly captured by these two models. XGBoost and SRF tend to outperform RF and SVR with higher KGE and CC (seen in Table 3). SRF produced the highest CC for each year ranging from 0.68 to 0.95, and produced the lowest MAE and RMSE for most validation years, ranging from 126.67 mm to 353.75 mm and 153.59 mm to 403.48 mm, respectively, which implied its good capability in fitting the non-linear relationships. Regarding KGE, XGBoost obtained the highest values for all years from 0.46 to 0.79.

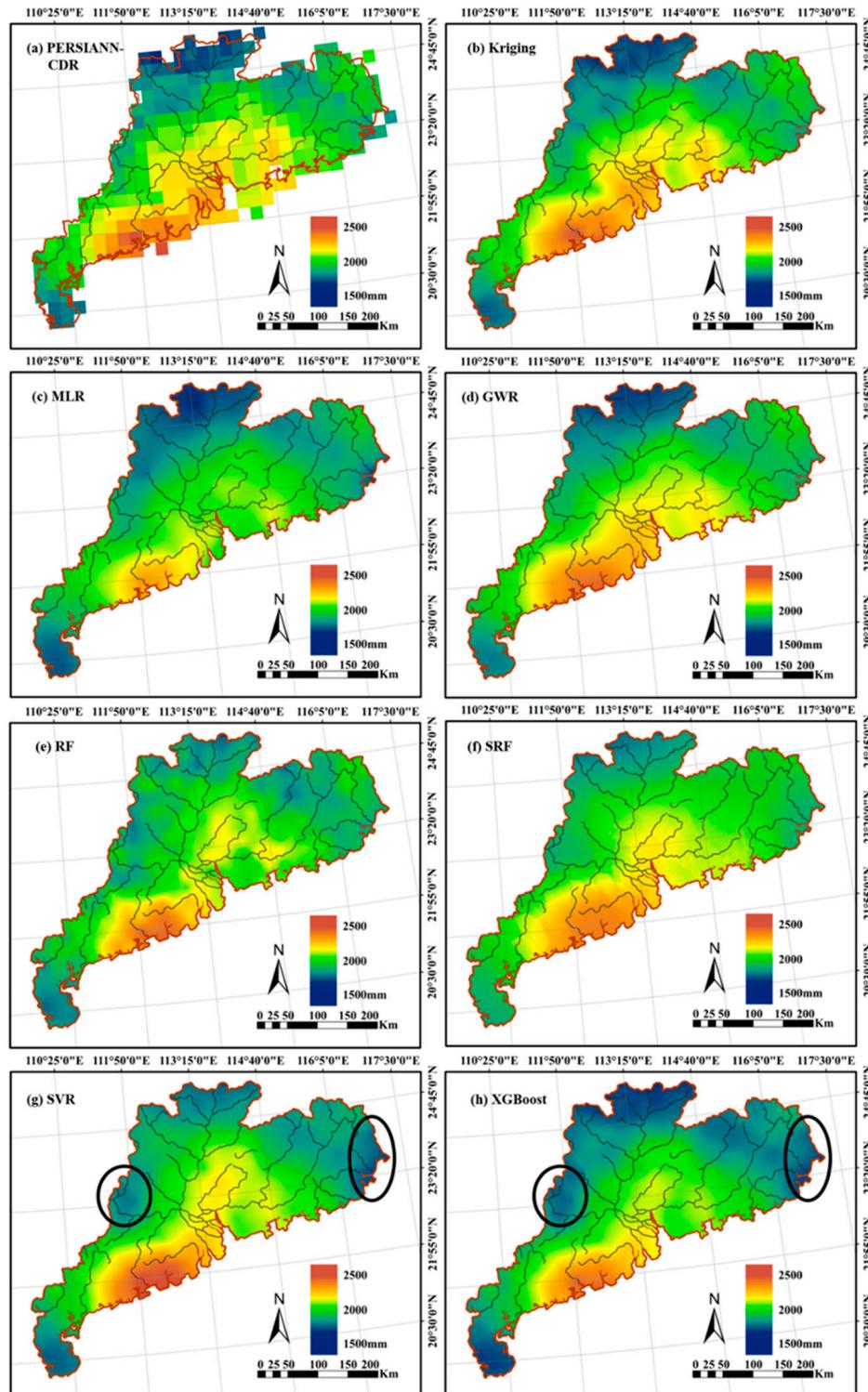
**Table 3.** Validation results of downscaling approaches for the testing period.

	Dataset	2006	2007	2008	2009	2010
CC	PERSIANN-CDR	0.52	0.47	0.60	0.79	0.51
	Kriging	0.57	0.51	0.56	0.80	0.48
	MLR	0.63	0.55	0.68	0.84	0.53
	GWR	0.66	0.62	0.73	0.89	0.49
	RF	0.78	0.69	0.82	0.91	0.64
	SRF	0.79	0.71	0.84	0.95	0.62
	SVR	0.77	0.65	0.79	0.90	0.68
	XGBoost	0.78	0.71	0.80	0.93	0.62
	PERSIANN-CDR	351.99	247.20	367.76	244.33	357.02
	Kriging	341.71	240.34	362.81	239.91	355.26
MAE (mm)	MLR	313.69	241.83	326.73	267.97	282.81
	GWR	278.62	203.05	281.33	184.33	350.51
	RF	244.63	193.28	222.57	172.37	299.61
	SRF	236.56	186.56	211.41	126.67	353.75
	SVR	245.48	196.61	243.70	184.19	302.78
	XGBoost	244.43	188.79	247.21	139.59	270.83
	PERSIANN-CDR	435.96	294.62	446.05	297.34	415.81
	Kriging	426.48	291.53	444.10	289.80	414.34
RMSE (mm)	MLR	417.41	291.33	415.81	315.71	332.88
	GWR	355.34	249.63	353.08	223.15	409.47
	RF	308.46	233.58	282.48	205.04	355.91
	SRF	298.42	224.72	268.24	153.89	403.48
	SVR	310.61	245.30	305.09	217.24	353.59
	XGBoost	305.13	226.37	309.09	171.29	318.29
	PERSIANN-CDR	0.36	0.35	0.57	0.67	0.35
	Kriging	0.38	0.36	0.54	0.70	0.31
KGE	MLR	0.32	0.32	0.57	0.59	0.30
	GWR	0.48	0.45	0.69	0.75	0.31
	RF	0.51	0.46	0.72	0.72	0.35
	SRF	0.56	0.50	0.73	0.79	0.34
	SVR	0.51	0.42	0.71	0.72	0.49
	XGBoost	0.57	0.53	0.73	0.79	0.46

### 3.2. Spatial Distribution of the Downscaled Results

Figure 3 shows the spatial distribution patterns of the original PERSIANN-CDR and downscaled results in the year 2010. All of the downscaled precipitation maps show similar distribution patterns to the original PERSIANN-CDR map, where there was much higher precipitation in the middle and lower precipitation in other areas. This is not a surprise, as all of the regression models were trained from satellite precipitation and would contain similar distribution characteristics as in the PERSIANN-CDR map. The annual precipitation map of the original PERSIANN-CDR contained some mosaic-like pixels due to the coarse resolution, while the downscaled maps of the ML-based algorithms provided enriched spatial information and reproduced basic spatial features. GWR worked well not only in terms of accuracy but also in capturing the spatial features of the PERSIANN-CDR distribution. MLR generally obtained higher CC and KGE than Kriging, but it failed to reproduce the spatial distribution and underestimated the precipitation in the middle area. Kriging and the original PERSIANN-CDR both had poor results of CC and KGE, and they produced almost the same spatial pattern, which might because Kriging interpolation could only generate the smooth values of the original satellite data. In Figure 3, SVR and XGBoost provided more details with large spatial variations. For example, in the regions highlighted

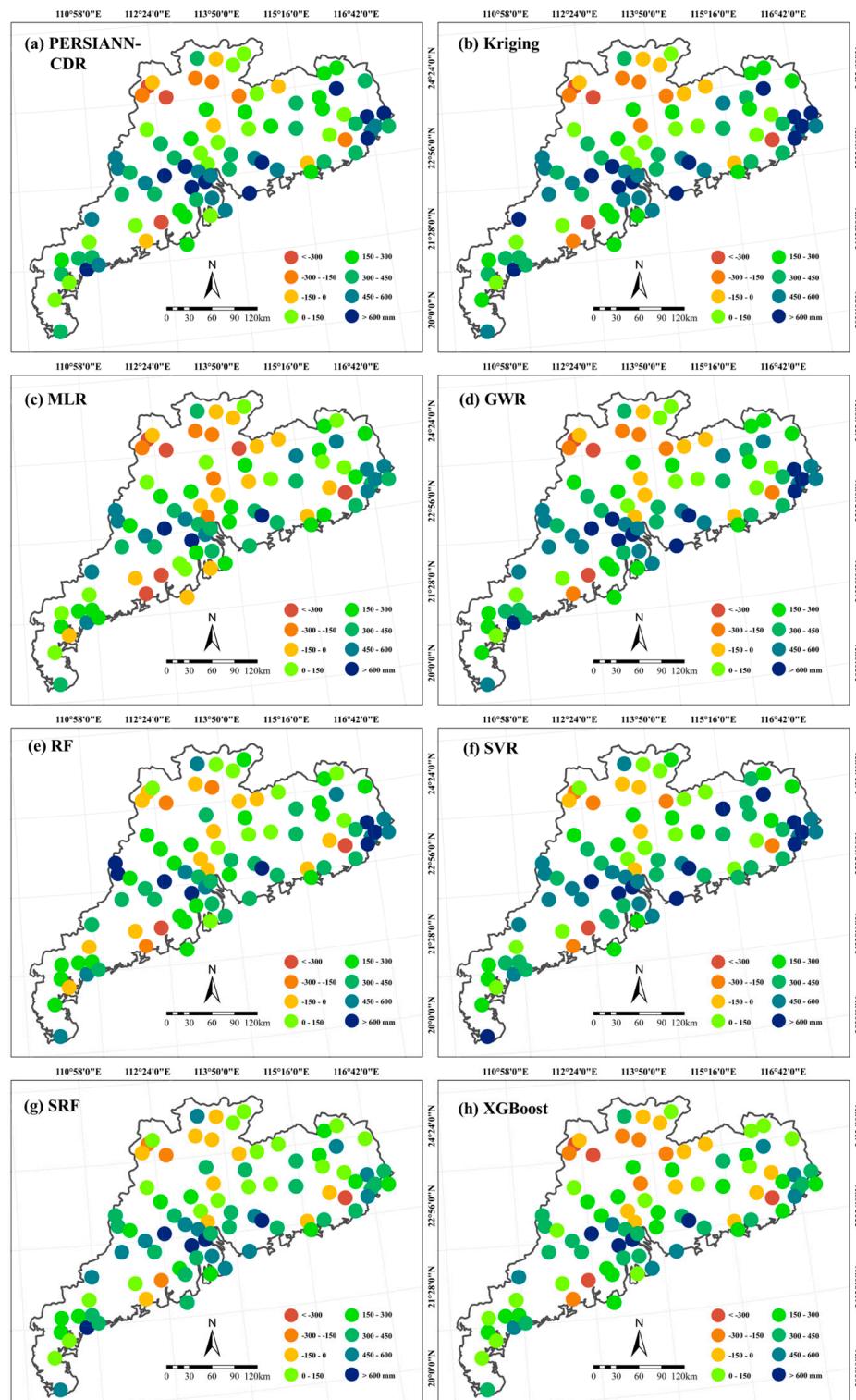
by the black circles (Figure 3g,h), the downscaled maps of SVR and XGBoost reproduced the low precipitation in these two areas.



**Figure 3.** Downscaled precipitation maps based on different methods for annual precipitation of 2010 over Guangdong Province.

The spatial distribution of the prediction bias of downscaling results was analysed based on the individual rain gauges in the year 2010. As shown in Figure 4, PERSIANN-CDR tended to overestimate at most rain gauges, especially in the middle and eastern areas where the annual precipitation was higher, indicating that PERSIANN-CDR had difficulty

in estimating heavy rainfall values. Similarly, other downscaled results reproduced this spatial pattern, as both the over- and under-estimation involved in the PERSIANN-CDR data were conclusively brought to the downscaled results. The number of rain gauges with overestimation in the results of the four ML methods (Figure 4e–h) decreased compared with the other models.



**Figure 4.** Spatial distribution of the prediction bias of the downscaling methods for annual precipitation of 2010 over Guangdong Province.

### 3.3. Were Residual Correlation and Calibration Procedures Helpful?

To explore whether residual correction helped to improve the precipitation estimation, the accuracy of the downscaling results based on the four ML methods before and after residual correction (termed RF\_RC, SRF\_RC, SVR\_RC and XGBoost\_RC) is summarised in Table 4. The residual correction process did not improve the precipitation estimation of the four ML-based methods. It undermined the CC and KGE values, and increased the MAE and RMSE. The partial results after residual correction even performed worse than the original PERSIANN-CDR. Table 5 shows the results of KGE, CC, MAE and RMSE after the GDA calibration framework, and the validation results are based on the observations from half of the rain gauges. The GDA calibration improved the accuracy of the precipitation estimation, and it significantly decreased the MAE and RMSE. RF\_GDA and SVR\_GDA received greater improvements than XGBoost\_GDA and SRF\_GDA, with KGE ranging from 0.41 to 0.86 and from 0.54 to 0.84 mm, respectively. The KGEs of precipitation estimation before and after residual correction based on different ML algorithms are presented in Figure 5. The results after residual correction and calibration were affected by the downscaled results directly. The precipitation data after residual correction generated the smallest accuracy. The results of the ML-based methods after GDA calibration achieved the best performance.

**Table 4.** Performance of the results after residual correction. The satellite precipitation was first downscaled using different ML-based methods, and then the downscaled results were under residual correction, which were termed RF\_RC, SRF\_RC, SVR\_RC and XGBoost\_RC.

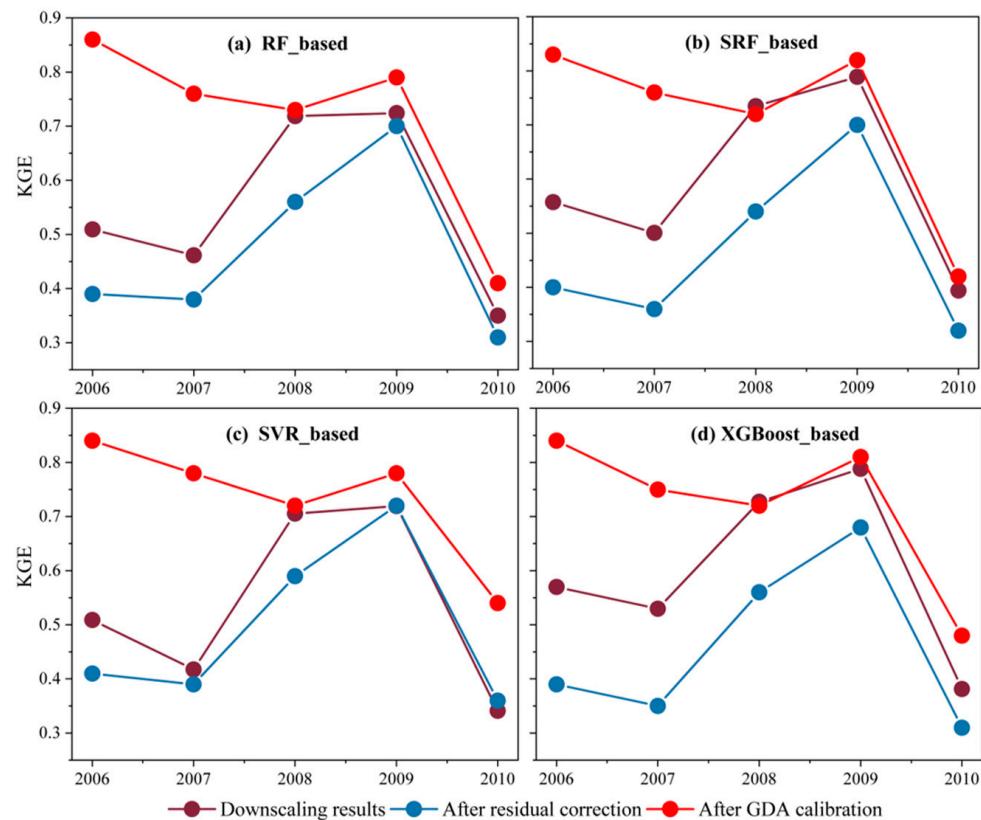
	Datasets	2006	2007	2008	2009	2010
CC	RF_RC	0.60	0.57	0.59	0.84	0.48
	SRF_RC	0.59	0.52	0.56	0.81	0.48
	SVR_RC	0.61	0.56	0.62	0.84	0.60
	XGBoost_RC	0.58	0.50	0.59	0.81	0.49
MAE (mm)	RF_RC	331.28	231.78	349.56	221.55	354.22
	SRF_RC	336.53	239.07	362.81	236.6	353.68
	SVR_RC	330.26	232.75	342.11	226.41	355.46
	XGBoost_RC	338.54	245.87	351.14	224.49	354.98
RMSE (mm)	RF_RC	414.38	279.13	426.51	265.11	413.86
	SRF_RC	419.57	289.92	444.09	285.15	412.52
	SVR_RC	410.27	284.03	420.4	269.3	407.87
	XGBoost_RC	424.13	293.98	429.11	272.15	414.12
KGE	RF_RC	0.39	0.38	0.56	0.70	0.31
	SRF_RC	0.40	0.36	0.54	0.70	0.32
	SVR_RC	0.41	0.39	0.59	0.72	0.36
	XGBoost_RC	0.39	0.35	0.56	0.68	0.31

Figure 6 presents maps of three different results in the validation year of 2010: (1) downscaled precipitation based on four ML methods, (2) downscaled precipitation after residual correction and (3) downscaled precipitation after calibration in the validation year of 2010. In the GDA calibration, observations from half of the rain gauges were used for merging (Figure 6b), and the other half were applied for the validation (Figure 6c). The results of the four ML-based methods showed different spatial patterns before residual correction, whereas the final results after residual correction had similar spatial patterns which were consistent with those of the original PERSIANN-CDR. Additionally, precipitation after the residual correction declined in the northern side and increased in the east compared with the results before residual correction. The results after calibration provided larger spatial variability in precipitation patterns and generated more spatial details compared with those before calibration. For example, the results after GDA calibration produced low precipitation values in the regions of the red circles in Figure 6, while the results after residual correction did not reproduce this feature. For the middle region, the downscaled

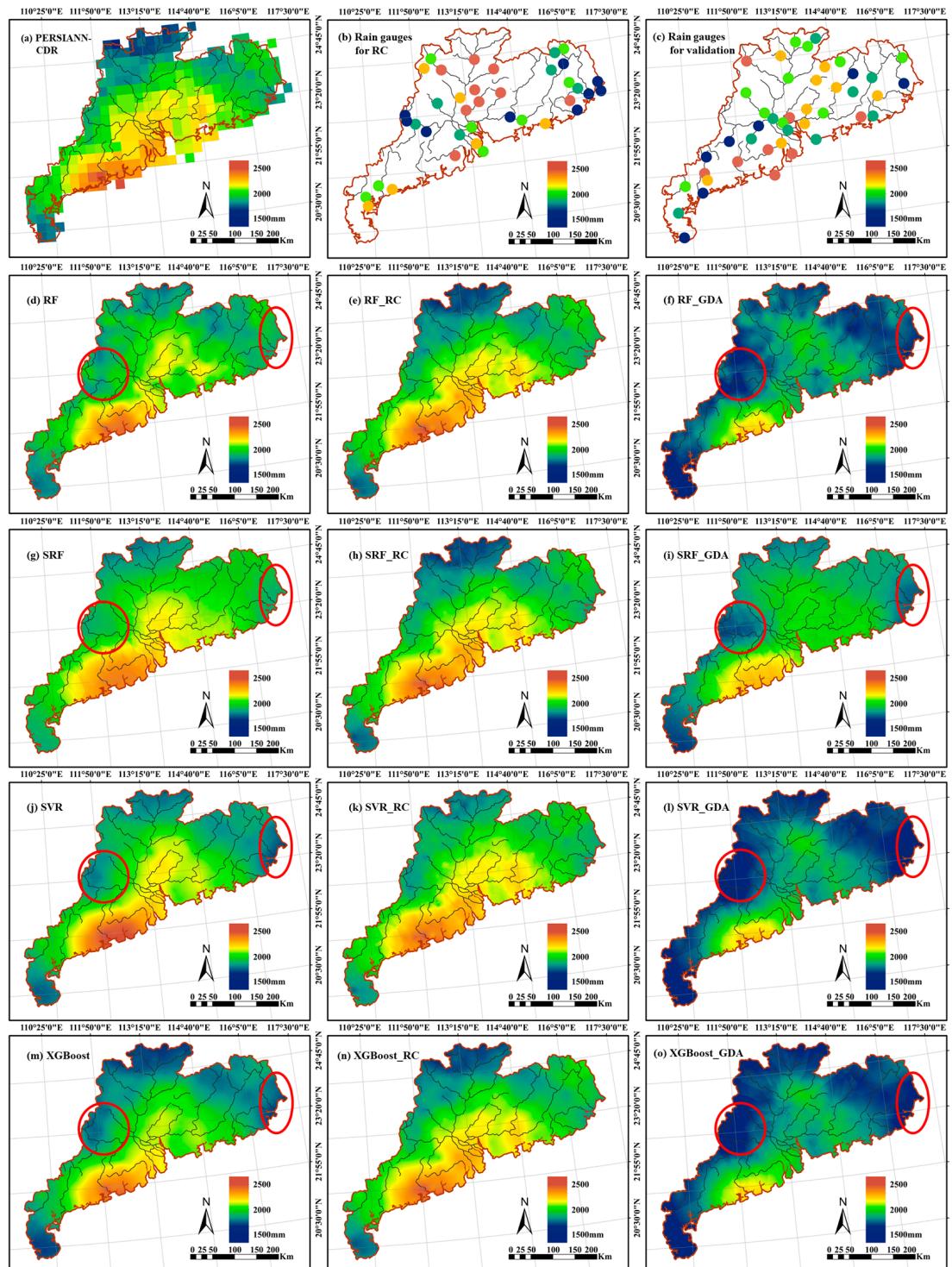
results and results after residual correction generated high precipitation over a large area, but the results after GDA calibration relieved the overestimation in this region.

**Table 5.** Accuracy of the results after GDA calibration. The downscaling results were then calibrated using GDA and the datasets obtained were interpreted as RF\_GDA, SRF\_GDA, SVR\_GDA and XGBoost\_GDA.

	Datasets	2006	2007	2008	2009	2010
CC	RF_GDA	0.91	0.78	0.82	0.93	0.69
	SRF_GDA	0.89	0.79	0.81	0.94	0.74
	SVR_GDA	0.90	0.79	0.78	0.95	0.70
	XGBoost_GDA	0.92	0.79	0.80	0.92	0.73
MAE (mm)	RF_GDA	154.93	146.49	197.07	119.18	186.82
	SRF_GDA	165.45	147.84	202.16	108.39	231.86
	SVR_GDA	165.72	146.06	209.06	109.58	187.47
	XGBoost_GDA	153.7	143.32	204.15	125.92	176.12
RMSE (mm)	RF_GDA	209.42	189.83	238.7	154.87	238.21
	SRF_GDA	220.02	189.93	246.94	143.5	272.21
	SVR_GDA	215.21	192.97	257.62	137.81	235.53
	XGBoost_GDA	194.00	183.27	246.67	159.09	224.19
KGE	RF_GDA	0.86	0.76	0.73	0.79	0.41
	SRF_GDA	0.83	0.76	0.72	0.82	0.42
	SVR_GDA	0.84	0.75	0.72	0.81	0.48
	XGBoost_GDA	0.84	0.78	0.72	0.78	0.54



**Figure 5.** Comparison of the KGE of four ML-based downscaling results, results after residual correction and results after calibration.



**Figure 6.** Spatial distribution of original PERSIANN-CDR; the rain gauges used in GDA and validation are shown in (a–c). The spatial distribution of residual correction and calibration results for annual precipitation of 2010 over Guangdong Province are shown in (d–o).

## 4. Discussion

### 4.1. Downscaled Results Based on ML Methods

In this study, the performance of four ML algorithms and three conventional downscaling models was evaluated, with ML algorithms generally outperforming the others. This can be attributed to two reasons. First, the ML models were better at capturing the non-linear relationships between environmental predictors and the precipitation, and were

more robust to outliers and noise in the data, which traditional downscaling models cannot handle as effectively [36]. Second, ML can automatically select the most important features relevant to precipitation, whereas conventional models rely on manual feature selection [56,59]. This can result in a more efficient and effective training process, improving the accuracy and robustness of the model and avoiding overfitting in the modelling [66].

Among the ML-based methods, XGBoost had the best KGE results, and outperformed the SRF, SVR and RF models. This could be attributed to the boosting strategy used in XGBoost, which helped to reduce the variance by aggregating the predictions of multiple weaker models and avoid overfitting by using regularisation techniques, such as tree pruning [25]. The better performance of SRF over RF and SVR might be due to incorporating spatial information into the ML model, and thus it had improved predictive accuracy compared to the SVR and RF models [9,55]. Similarly, GWR achieved better results than MLR by implementing geographical weighting to the features in each of the local regression equations [22,67]. These results highlighted the importance of spatial information in downscaling models. However, SRF still received slightly worse downscaling results than XGBoost, possibly because the spatial heterogeneity of precipitation was not well captured by the spatial random forest model [9,27], and XGBoost had many hyperparameters that can be tuned to optimise the model's performance [18]. SVR had better performance than RF, which might be attributed to the fact that the SVR model was less prone to overfitting and had better generalisation ability than RF [34,56].

This study found that ML algorithms were more resilient than conventional downscaling methods, and XGBoost achieved the highest performance out of the four ML algorithms. However, the performance of the downscaling results varied not only with different algorithms, but also with different years. As shown in Table 3 and Figure 6, the KGE and CC of the downscaling results in 2009 had a higher value, but the results of 2007 and 2010 were lower. The poor estimation in these years might be because the satellite data could not be perfectly modelled by DEM, NDVI and LST. Meanwhile, the capability of models to explain the satellite precipitation and environmental factors could impact the downscaling performance, but more importantly, the accuracy of the original PERSIANN-CDR could be a main error source in the downscaling results [68]. The error source analysis was not contained in this study, but our results supported the findings in previous studies [20–22] that the downscaling accuracy was largely determined by the quality of the original satellite data.

#### 4.2. Performance of Downscaling Results after GDA Calibration and Residual Correlation

The downscaled precipitation showed more accurate results after GDA calibration, with increased CC and KGE, as well as decreased MAE and RMSE. The improved performance after GDA calibration was consistent with the previous studies [9,34,38,49]. To further investigate whether the effect of GDA on improving accuracy was caused by the downscaling process, a comparative experiment of directly performing the GDA on the original satellite data was conducted. Table 6 shows the accuracy of satellite precipitation data after GDA calibration. As shown, the GDA process could also improve the accuracy of the original satellite data, but compared with the downscaled results after calibration in Table 5, the improvement was not as significant as that achieved based on the downscaled data, which suggests the importance of the downscaling process before the calibration. This might be attributed to the fact that the downscaled precipitation had higher spatial resolution, and could strengthen the performance of the calibration. In the calibration procedure, the precipitation between the pixels and rain gauges was compared. However, the precipitation in each pixel represented the areal average precipitation within it, while the rain gauge measurements were point-based. The mismatch between the gridded and the point-based values might have affected the calibration results. Downscaling provided finer-resolution gridded data, and allowed for a more effective comparison with the rain gauge data.

**Table 6.** Accuracy of the original PERSIANN-CDR after GDA calibration for each year.

Year	R	MAE	RMSE	KGE
2006	0.80	199.30	262.72	0.75
2007	0.73	173.37	220.48	0.61
2008	0.68	239.28	310.46	0.65
2009	0.87	163.28	216.34	0.73
2010	0.64	249.46	304.56	0.36

For the residual correction, this study found that the estimated precipitation after residual correction was poorer than before, which was consistent with the results of previous studies using the GWR model [22,59]. However, some studies reported opposite results, indicating an improvement in precipitation data after residual correction. Zhao [69] illustrated that residual correction played an important role in RF-based downscaling. Residual correction greatly increased the accuracy of the precipitation estimation conducted by artificial neural networks [33]. The worse performance after residual correction in our study might be explained by the following reasons: (1) The uncertainty in the original satellite data was inherited in the results after residual correction. The residuals were the difference between the estimation from models and the original PERSIANN-CDR with coarse resolution, and reflected the precipitation that could not be predicted by the models. However, if the original satellite data had significant errors, it would be difficult for residual correlation to produce accurate results. (2) The residuals with coarse resolution were interpolated to 1 km using the ordinary Kriging method in this study, but other studies have adopted different methods including spline, IDW and nearest neighbour interpolation methods. The difference between the performance of Kriging and other interpolation techniques has not been investigated sufficiently, which might affect the application of residual correction [2,70]. (3) The interpolation of residuals using the Kriging method only took the distance and direction into account and did not introduce other environmental variables. The fewer predictors used in residual correlation may not be appropriate and they cannot be used to capture the underlying patterns in the residuals, which could lead to incorrect adjustments and worse results.

Furthermore, quantitative analysis of the effect of residual correction was implemented to better understand why the accuracy decreased after residual correction. Noting that the error of the original satellite precipitation and the estimated precipitation at 25 km were  $\epsilon(S_L)$  and  $\epsilon(D_L)$ , the error of the residuals at 25 km would be  $|\epsilon(S_L) - \epsilon(D_L)|$ . The error of the downscaled precipitation with a high resolution was  $\epsilon(D_H)^2$ . Incorporating the residual correction, the error of the precipitation would be  $[\epsilon(S_L) - \epsilon(D_L) + \epsilon(D_H)]^2$ . The impact of the residual correction depends on the value of the error before and after the residual correction, i.e., the value of  $\epsilon(D_H)^2$  and  $[\epsilon(S_L) - \epsilon(D_L) + \epsilon(D_H)]^2$ . Specifically, the residual correction yields a positive effect when the expression 4.1 is negative. Accordingly, when the  $\epsilon(S_L)$  is smaller than the  $\epsilon(D_L)$ , or the  $\epsilon(S_L)$  is large enough, the residual correction process will decrease the accuracy of the precipitation. Therefore, when the accuracy of the original satellite data is not good, the residual correction should not be conducted after downscaling.

$$\begin{aligned} & [\epsilon(S_L) - \epsilon(D_L) + \epsilon(D_H)]^2 - \epsilon(D_H)^2 \\ &= [\epsilon(S_L) - \epsilon(D_L)] * [\epsilon(S_L) - \epsilon(D_L) + 2\epsilon(D_H)] \end{aligned} \quad (1)$$

#### 4.3. Future Perspectives

This study demonstrated that ML-based approaches had more potential to be applied in precipitation downscaling compared with other conventional downscaling methods, and the XGBoost outperformed the others in generating high-resolution precipitation data. However, there is still space for improvement in further studies. First, a larger sample size of training datasets is required to construct the ML downscaling models, as the limited data used in this study might lead to overfitting and poor generalisation of the ML models.

Generally, the size and diversity of training datasets have a significant impact on the performance of machine learning models, and a larger and high-quality training dataset can help to improve the accuracy, generalisation ability and robustness of ML models [71,72]. Secondly, more observations from rain gauges should be involved for the testing of ML models. In this study, the validation results were based on in situ data from only 86 rain gauge stations. The limited spatial representativeness with only a few rain gauges made it difficult to correctly capture the spatial variability in precipitation across the entire region, which could lead to incomplete and potentially biased results [27]. Thirdly, multiple satellite-derived precipitation data sources would be necessary and meaningful in further downscaling studies. The PERSIANN-CDR was used as the original satellite-derived data in this study to develop the ML-based downscaling models due to its longer temporal coverage and good consistency with measurements. However, there are many other remote sensing precipitation products, each with their limitations and advantages. To reduce the uncertainty contained in satellite data, the combination of multiple satellite-derived precipitation data would provide a more reliable estimation than individual precipitation products [37]. Furthermore, with the advent of the “big data” era, deep learning has shown great potential in the downscaling of precipitation in recent studies [73]. The application of many deep learning algorithms, such as convolutional neural networks (CNNs) [74], long short-term memory (LSTM) networks [75] and generative adversarial network (GAN) [76], has displayed high accuracy and efficiency of precipitation downscaling in comparison to traditional regression-based models. Evaluating the performance of different deep learning models is beyond the scope of this study, but it could be explored in our future work.

## 5. Conclusions

This study aimed to present an accurate and reliable downscaling scheme for satellite-based precipitation data. The performance of four ML algorithms and three conventional downscaling models was evaluated, and residual correction as well as calibration procedures were implemented to the downscaled results to examine possible improvements. The results showed that the ML-based methods worked better in producing high-resolution precipitation estimation than the classical interpolation methods. Specifically, XGBoost achieved the best results in downscaling, followed by the SRF, SVR and RF algorithms. The GDA calibration process significantly improved the downscaled results and is an essential step in the downscaling scheme. However, no significant improvement was found after the residual correction, indicating that residual correction should not be included in the ML-based downscaling framework. Therefore, for an accurate and reliable downscaling scheme, XGBoost should be used as the primary downscaling method, followed by the calibration-based post-processing of the downscaled results. This downscaling scheme demonstrated the most optimal downscaled results, and it can be applied in regions where high-resolution precipitation is consistently required. In our future work, more emphasis could be placed on downscaling with a temporal resolution, such as on the daily and hourly scales.

**Author Contributions:** Conceptualisation, H.Z. and Q.Z.; methodology, H.Z., H.L. and Q.Z.; supervision, Q.Z.; validation, H.Z. and A.C.; formal analysis, H.Z. and A.C.; writing—original draft preparation, H.Z.; writing—review and editing, H.Z., H.L. and Q.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** The research is supported by the Natural Science Foundation of China (NSFC) General Program (41971386 and 42271416) and Hong Kong Research Grant Council (RGC) General Research Fund (HKBU 12301820).

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author.

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

## References

- Shi, Y.; Song, L.; Xia, Z.; Lin, Y.; Myneni, R.B.; Choi, S.; Wang, L.; Ni, X.; Lao, C.; Yang, F. Mapping annual precipitation across Mainland China in the period 2001–2010 from TRMM3B43 product using spatial downscaling approach. *Remote Sens.* **2015**, *7*, 5849. [[CrossRef](#)]
- Jing, W.; Yang, Y.; Yue, X.; Zhao, X. A comparison of different regression algorithms for downscaling monthly satellite-based precipitation over North China. *Remote Sens.* **2016**, *8*, 835. [[CrossRef](#)]
- Guo, H.; Bao, A.; Liu, T.; Chen, S.; Ndayisaba, F. Evaluation of PERSIANN-CDR for meteorological drought monitoring over China. *Remote Sens.* **2016**, *8*, 379. [[CrossRef](#)]
- Zhang, J.; Xu, J.; Dai, X.; Ruan, H.; Liu, X.; Jing, W. Multi-Source Precipitation Data Merging for Heavy Rainfall Events Based on Cokriging and Machine Learning Methods. *Remote Sens.* **2022**, *14*, 1750. [[CrossRef](#)]
- Chu, H.J.; Wijayanti, R.F.; Jaelani, L.M.; Tsai, H.P. Time varying spatial downscaling of satellite-based drought index. *Remote Sens.* **2021**, *13*, 3693. [[CrossRef](#)]
- Elnashar, A.; Zeng, H.; Wu, B.; Zhang, N.; Tian, F.; Zhang, M.; Zhu, W.; Yan, N.; Chen, Z.; Sun, Z.; et al. Downscaling TRMM monthly precipitation using google earth engine and google cloud computing. *Remote Sens.* **2020**, *12*, 3860. [[CrossRef](#)]
- Fan, D.; Wu, H.; Dong, G.; Jiang, X.; Xue, H. A temporal disaggregation approach for TRMM monthly precipitation products using AMSR2 soil moisture data. *Remote Sens.* **2019**, *11*, 2962. [[CrossRef](#)]
- Funk, C.; Peterson, P.; Landsfeld, M.; Pedreros, D.; Verdin, J.; Shukla, S.; Husak, G.; Rowland, J.; Harrison, L.; Hoell, A.; et al. The climate hazards infrared precipitation with stations—A new environmental record for monitoring extremes. *Sci. Data* **2015**, *2*, 1–21. [[CrossRef](#)]
- Chen, C.; Hu, B.; Li, Y. Easy-to-use spatial Random Forest-based downscaling-calibration method for producing high resolution and accurate precipitation data. *Hydrol. Earth Syst. Sci.* **2021**, *25*, 5667–5682. [[CrossRef](#)]
- Kummerow, C.; Simpson, J.; Thiele, O.; Barnes, W.; Chang, A.T.C.; Stocker, E.; Adler, R.F.; Hou, A.; Kakar, R.; Wentz, F.; et al. The status of the tropical rainfall measuring mission (TRMM) after two years in orbit. *J. Appl. Meteorol.* **2000**, *39*, 1965–1982. [[CrossRef](#)]
- Kummerow, C.; Barnes, W.; Kozu, T.; Shiue, J.; Simpson, J. The Tropical Rainfall Measuring Mission (TRMM) sensor package. *J. Atmos. Ocean. Technol.* **1998**, *15*, 809–817. [[CrossRef](#)]
- Hou, A.Y.; Kakar, R.K.; Neeck, S.; Azarbarzin, A.A.; Kummerow, C.D.; Kojima, M.; Oki, R.; Nakamura, K.; Iguchi, T. The global precipitation measurement mission. *Bull. Am. Meteorol. Soc.* **2014**, *95*, 701–722. [[CrossRef](#)]
- Huffman, G.J.; Adler, R.F.; Arkin, P.; Chang, A.; Ferraro, R.; Gruber, A.; Janowiak, J.; McNab, A.; Rudolf, B.; Schneider, U. The Global Precipitation Climatology Project (GPCP) Combined Precipitation Dataset. *Bull. Am. Meteorol. Soc.* **1997**, *78*, 5–20. [[CrossRef](#)]
- Adler, R.F.; Huffman, G.J.; Chang, A.; Ferraro, R.; Xie, P.P.; Janowiak, J.; Rudolf, B.; Schneider, U.; Curtis, S.; Bolvin, D.; et al. The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979–present). *J. Hydrometeorol.* **2003**, *4*, 1147–1167. [[CrossRef](#)]
- Hsu, K.L.; Gao, X.; Sorooshian, S.; Gupta, H.V. Precipitation estimation from remotely sensed information using artificial neural networks. *J. Appl. Meteorol.* **1997**, *36*, 1176–1190. [[CrossRef](#)]
- Ashouri, H.; Hsu, K.L.; Sorooshian, S.; Braithwaite, D.K.; Knapp, K.R.; Cecil, L.D.; Nelson, B.R.; Prat, O.P. PERSIANN-CDR: Daily precipitation climate data record from multisatellite observations for hydrological and climate studies. *Bull. Am. Meteorol. Soc.* **2015**, *96*, 69–83. [[CrossRef](#)]
- Miao, C.; Ashouri, H.; Hsu, K.L.; Sorooshian, S.; Duan, Q. Evaluation of the PERSIANN-CDR daily rainfall estimates in capturing the behavior of extreme precipitation events over China. *J. Hydrometeorol.* **2015**, *16*, 1387–1396. [[CrossRef](#)]
- Ali, S.; Khorrami, B.; Jehanzaib, M.; Tariq, A.; Ajmal, M.; Arshad, A.; Shafeeque, M.; Dilawar, A.; Basit, I.; Zhang, L.; et al. Spatial Downscaling of GRACE Data Based on XGBoost Model for Improved Understanding of Hydrological Droughts in the Indus Basin Irrigation System (IBIS). *Remote Sens.* **2023**, *15*, 873. [[CrossRef](#)]
- Abdollahipour, A.; Ahmadi, H.; Aminnejad, B. A review of downscaling methods of satellite-based precipitation estimates. *Earth Sci. Inform.* **2022**, *15*, 1–20. [[CrossRef](#)]
- Immerzeel, W.W.; Rutten, M.M.; Droogers, P. Spatial downscaling of TRMM precipitation using vegetative response on the Iberian Peninsula. *Remote Sens. Environ.* **2009**, *113*, 362–370. [[CrossRef](#)]
- Jia, S.; Zhu, W.; Lu, A.; Yan, T. A statistical spatial downscaling algorithm of TRMM precipitation based on NDVI and DEM in the Qaidam Basin of China. *Remote Sens. Environ.* **2011**, *115*, 3069–3079. [[CrossRef](#)]
- Xu, S.; Wu, C.; Wang, L.; Gonsamo, A.; Shen, Y.; Niu, Z. A new satellite-based monthly precipitation downscaling algorithm with non-stationary relationship between precipitation and land surface characteristics. *Remote Sens. Environ.* **2015**, *162*, 119–140. [[CrossRef](#)]
- Shirali, E.; Nikbakht Shahbazi, A.; Fathian, H.; Zohrabi, N.; Mobarak Hassan, E. Evaluation of WRF and artificial intelligence models in short-term rainfall, temperature and flood forecast (case study). *J. Earth Syst. Sci.* **2020**, *129*, 188. [[CrossRef](#)]
- Baghanam, A.H.; Eslahi, M.; Sheikhabaei, A.; Seifi, A.J. Assessing the impact of climate change over the northwest of Iran: An overview of statistical downscaling methods. *Theor. Appl. Climatol.* **2020**, *141*, 1135–1150. [[CrossRef](#)]
- Liu, H.; Li, Q.; Bai, Y.; Yang, C.; Wang, J.; Zhou, Q.; Hu, S.; Shi, T.; Liao, X.; Wu, G. Improving satellite retrieval of oceanic particulate organic carbon concentrations using machine learning methods. *Remote Sens. Environ.* **2021**, *256*, 112316. [[CrossRef](#)]

26. Chen, S.T.; Yu, P.S.; Tang, Y.H. Statistical downscaling of daily precipitation using support vector machines and multivariate analysis. *J. Hydrol.* **2010**, *385*, 13–22. [[CrossRef](#)]
27. He, X.; Chaney, N.W.; Schleiss, M.; Sheffield, J. Spatial downscaling of precipitation using adaptable random forests. *Water Resour. Res.* **2016**, *52*, 8217–8237. [[CrossRef](#)]
28. Devak, M.; Dhanya, C.T.; Gosain, A.K. Dynamic coupling of support vector machine and K-nearest neighbour for downscaling daily rainfall. *J. Hydrol.* **2015**, *525*, 286–301. [[CrossRef](#)]
29. Yan, X.; Chen, H.; Tian, B.; Sheng, S.; Wang, J.; Kim, J.S. A downscaling–merging scheme for improving daily spatial precipitation estimates based on random forest and cokriging. *Remote Sens.* **2021**, *13*, 2040. [[CrossRef](#)]
30. Pour, S.H.; Shahid, S.; Chung, E.S. A Hybrid Model for Statistical Downscaling of Daily Rainfall. *Procedia Eng.* **2016**, *154*, 1424–1430. [[CrossRef](#)]
31. Raje, D.; Mujumdar, P.P. A comparison of three methods for downscaling daily precipitation in the Punjab region. *Hydrol. Process.* **2011**, *25*, 3575–3589. [[CrossRef](#)]
32. Nasseri, M.; Tavakol-Davani, H.; Zahraie, B. Performance assessment of different data mining methods in statistical downscaling of daily precipitation. *J. Hydrol.* **2013**, *492*, 1–14. [[CrossRef](#)]
33. Sharifi, E.; Saghafian, B.; Steinacker, R. Downscaling Satellite Precipitation Estimates With Multiple Linear Regression, Artificial Neural Networks, and Spline Interpolation Techniques. *J. Geophys. Res. Atmos.* **2019**, *124*, 789–805. [[CrossRef](#)]
34. Karbalaye Ghorbanpour, A.; Hessels, T.; Moghim, S.; Afshar, A. Comparison and assessment of spatial downscaling methods for enhancing the accuracy of satellite-based precipitation over Lake Urmia Basin. *J. Hydrol.* **2021**, *596*, 126055. [[CrossRef](#)]
35. Sahour, H.; Sultan, M.; Vazifedan, M.; Abdelmohsen, K.; Karki, S.; Yellich, J.A.; Gebremichael, E.; Alshehri, F.; Elbayoumi, T.M. Statistical applications to downscale GRACE-derived terrestrialwater storage data and to fill temporal gaps. *Remote Sens.* **2020**, *12*, 533. [[CrossRef](#)]
36. Zhang, J.; Liu, K.; Wang, M. Downscaling groundwater storage data in China to a 1-km resolution using machine learning methods. *Remote Sens.* **2021**, *13*, 523. [[CrossRef](#)]
37. Ulloa, J.; Ballari, D.; Campozano, L.; Samaniego, E. Two-step downscaling of TRMM 3b43 V7 precipitation in contrasting climatic regions with sparse monitoring: The case of Ecuador in tropical South America. *Remote Sens.* **2017**, *9*, 758. [[CrossRef](#)]
38. Duan, Z.; Bastiaanssen, W.G.M. First results from Version 7 TRMM 3B43 precipitation product in combination with a new downscaling-calibration procedure. *Remote Sens. Environ.* **2013**, *131*, 1–13. [[CrossRef](#)]
39. Cheema, M.J.M.; Bastiaanssen, W.G.M. Local calibration of remotely sensed rainfall from the TRMM satellite for different periods and spatial scales in the Indus Basin. *Int. J. Remote Sens.* **2012**, *33*, 2603–2627. [[CrossRef](#)]
40. Xin, Y.; Lu, N.; Jiang, H.; Liu, Y.; Yao, L. Performance of ERA5 reanalysis precipitation products in the Guangdong-Hong Kong-Macao greater Bay Area, China. *J. Hydrol.* **2021**, *602*, 126791. [[CrossRef](#)]
41. Yan, M.; Chan, J.C.L.; Zhao, K. Impacts of Urbanization on the Precipitation Characteristics in Guangdong Province, China. *Adv. Atmos. Sci.* **2020**, *37*, 696–706. [[CrossRef](#)]
42. CMDSC China Meteorological Data Service Center: Gauge Data [Data Set]. Available online: <http://data.cma.cn/data/detail/dataCode/> (accessed on 1 December 2021).
43. CGIAR SRTM Data, CGIAR [Data Set]. Available online: [https://www.usgs.gov/centers/eros/science/usgs-eros-archive-digital-elevation-shuttle-radar-topography-mission-srtm-1?qt-science\\_center\\_objects=0#qt-science\\_center\\_objects](https://www.usgs.gov/centers/eros/science/usgs-eros-archive-digital-elevation-shuttle-radar-topography-mission-srtm-1?qt-science_center_objects=0#qt-science_center_objects) (accessed on 1 December 2021).
44. Shortridge, A.; Messina, J. Spatial structure and landscape associations of SRTM error. *Remote Sens. Environ.* **2011**, *115*, 1576–1587. [[CrossRef](#)]
45. National Tibetan Plateau/Third Pole Environment Data Center Global GIMMS NDVI3g v1 Dataset (1981–2015). 2018. Available online: <https://climatedataguide.ucar.edu/collections/climate-data-record> (accessed on 1 December 2021).
46. Tucker, C.J.; Pinzon, J.E.; Brown, M.E.; Slayback, D.A.; Pak, E.W.; Mahoney, R.; Vermote, E.F.; El Saleous, N. An extended AVHRR 8-km NDVI dataset compatible with MODIS and SPOT vegetation NDVI data. *Int. J. Remote Sens.* **2005**, *26*, 4485–4498. [[CrossRef](#)]
47. Pinzon, J.E.; Tucker, C.J. A non-stationary 1981–2012 AVHRR NDVI3g time series. *Remote Sens.* **2014**, *6*, 6929. [[CrossRef](#)]
48. Wan, Z.; Hook, S.; Hulley, G. MOD11A2 MODIS/Terra Land Surface Temperature/Emissivity 8-Day L3 Global 1 km SIN Grid V006 [Data set]. Nasa Eosdis Land Processes Daac. 2015. Available online: <https://ladsweb.modaps.eosdis.nasa.gov/missions-and-measurements/products/MOD11A2> (accessed on 1 December 2021).
49. Arshad, A.; Zhang, W.; Zhang, Z.; Wang, S.; Zhang, B.; Cheema, M.J.M.; Shalamzari, M.J. Reconstructing high-resolution gridded precipitation data using an improved downscaling approach over the high altitude mountain regions of Upper Indus Basin (UIB). *Sci. Total Environ.* **2021**, *784*, 147140. [[CrossRef](#)]
50. Zhang, Q.; Shen, Z.; Xu, C.Y.; Sun, P.; Hu, P.; He, C. A new statistical downscaling approach for global evaluation of the CMIP5 precipitation outputs: Model development and application. *Sci. Total Environ.* **2019**, *690*, 1048–1067. [[CrossRef](#)]
51. Zhang, T.; Li, B.; Yuan, Y.; Gao, X.; Sun, Q.; Xu, L.; Jiang, Y. Spatial downscaling of TRMM precipitation data considering the impacts of macro-geographical factors and local elevation in the Three-River Headwaters Region. *Remote Sens. Environ.* **2018**, *215*, 109–127. [[CrossRef](#)]
52. Zhan, C.; Han, J.; Hu, S.; Liu, L.; Dong, Y. Spatial Downscaling of GPM Annual and Monthly Precipitation Using Regression-Based Algorithms in a Mountainous Area. *Adv. Meteorol.* **2018**, *2018*, 1506017. [[CrossRef](#)]

53. Chai, Y.; Martins, G.; Nobre, C.; von Randow, C.; Chen, T.; Dolman, H. Constraining Amazonian land surface temperature sensitivity to precipitation and the probability of forest dieback. *Npj Clim. Atmos. Sci.* **2021**, *4*, 6. [[CrossRef](#)]
54. Shah, H.L.; Zhou, T.; Huang, M.; Mishra, V. Strong Influence of Irrigation on Water Budget and Land Surface Temperature in Indian Subcontinental River Basins. *J. Geophys. Res. Atmos.* **2019**, *124*, 1449–1462. [[CrossRef](#)]
55. Shen, Z.; Yong, B. Downscaling the GPM-based satellite precipitation retrievals using gradient boosting decision tree approach over Mainland China. *J. Hydrol.* **2021**, *602*, 126803. [[CrossRef](#)]
56. Hu, S.; Liu, H.; Zhao, W.; Shi, T.; Hu, Z.; Li, Q.; Wu, G. Comparison of machine learning techniques in inferring phytoplankton size classes. *Remote Sens.* **2018**, *10*, 191. [[CrossRef](#)]
57. Hengl, T.; Nussbaum, M.; Wright, M.N.; Heuvelink, G.B.M.; Gräler, B. Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ* **2018**, *2018*, e5518. [[CrossRef](#)]
58. Wright, M.N.; Ziegler, A. Ranger: A fast implementation of random forests for high dimensional data in C++ and R. *J. Stat. Softw.* **2017**, *77*. [[CrossRef](#)]
59. Sun, B.; Zhang, Y.; Zhou, Q.; Zhang, X. Effectiveness of Semi-Supervised Learning and Multi-Source Data in Detailed Urban Landuse Mapping with a Few Labeled Samples. *Remote Sens.* **2022**, *14*, 648. [[CrossRef](#)]
60. Dimitriadou, E.; Hornik, K.; Leisch, F.; Meyer, D.; Weingessel, A.; Leisch, M.F. Package ‘e1071’. R Software Package. 2009. Available online: <http://cran.r-project.org/web/packages/e1071/index.html> (accessed on 1 December 2021).
61. Cutler, A.; Cutler, D.R.; Stevens, J.R. Random forests. In *Ensemble Machine Learning: Methods and Applications*; Springer: New York, NY, USA, 2012; pp. 157–175.
62. Benito, B.M. spatialRF: Easy Spatial Regression with Random Forest. R Package Version 1.1.0. 2021. Available online: <https://blasbenito.github.io/spatialRF/> (accessed on 1 December 2021).
63. Chen, T.; He, T. Xgboost: Extreme Gradient Boosting. R Package Version 0.4-2 1.4. 2015, pp. 1–4. Available online: <https://cran.microsoft.com/snapshot/2017-12-11/web/packages/xgboost/vignettes/xgboost.pdf> (accessed on 1 December 2021).
64. Oshan, T. GWR4 2016. Available online: <https://gwrtools.github.io/gwr4-downloads.html> (accessed on 1 December 2021).
65. Kling, H.; Fuchs, M.; Paulin, M. Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *J. Hydrol.* **2012**, *424*–425, 264–277. [[CrossRef](#)]
66. Li, Q.; Wang, Z.; Shangguan, W.; Li, L.; Yao, Y.; Yu, F. Improved daily SMAP satellite soil moisture prediction over China using deep learning model with transfer learning. *J. Hydrol.* **2021**, *600*, 126698. [[CrossRef](#)]
67. Chen, Y.; Huang, J.; Sheng, S.; Mansaray, L.R.; Liu, Z.; Wu, H.; Wang, X. A new downscaling-integration framework for high-resolution monthly precipitation estimates: Combining rain gauge observations, satellite-derived precipitation data and geographical ancillary data. *Remote Sens. Environ.* **2018**, *214*, 154–172. [[CrossRef](#)]
68. Huang, X.; Swain, D.L.; Hall, A.D. Future precipitation increase from very high resolution ensemble downscaling of extreme atmospheric river storms in California. *Sci. Adv.* **2020**, *6*, eaba1323. [[CrossRef](#)]
69. Zhao, N. An efficient downscaling scheme for high-resolution precipitation estimates over a high mountainous watershed. *Remote Sens.* **2021**, *13*, 234. [[CrossRef](#)]
70. Jing, W.; Yang, Y.; Yue, X.; Zhao, X. A spatial downscaling algorithm for satellite-based precipitation over the Tibetan plateau based on NDVI, DEM, and land surface temperature. *Remote Sens.* **2016**, *8*, 655. [[CrossRef](#)]
71. Emadi, M.; Taghizadeh-Mehrjardi, R.; Cherati, A.; Danesh, M.; Mosavi, A.; Scholten, T. Predicting and mapping of soil organic carbon using machine learning algorithms in Northern Iran. *Remote Sens.* **2020**, *12*, 2234. [[CrossRef](#)]
72. Xu, X.; Du, C.; Ma, F.; Qiu, Z.; Zhou, J. A Framework for High-Resolution Mapping of Soil Organic Matter (SOM) by the Integration of Fourier Mid-Infrared Attenuation Total Reflectance Spectroscopy (FTIR-ATR), Sentinel-2 Images, and DEM Derivatives. *Remote Sens.* **2023**, *15*, 1072. [[CrossRef](#)]
73. Wang, F.; Tian, D.; Lowe, L.; Kalin, L.; Lehrter, J. Deep Learning for Daily Precipitation and Temperature Downscaling. *Water Resour. Res.* **2021**, *57*, e2020WR029308. [[CrossRef](#)]
74. Tu, T.; Ishida, K.; Ercan, A.; Kiyama, M.; Amagasaki, M.; Zhao, T. Hybrid precipitation downscaling over coastal watersheds in Japan using WRF and CNN. *J. Hydrol. Reg. Stud.* **2021**, *37*, 100921. [[CrossRef](#)]
75. Wu, H.; Yang, Q.; Liu, J.; Wang, G. A spatiotemporal deep fusion model for merging satellite and gauge precipitation in China. *J. Hydrol.* **2020**, *584*, 124664. [[CrossRef](#)]
76. Kumar, B.; Atey, K.; Singh, B.B.; Chattopadhyay, R.; Acharya, N.; Singh, M.; Nanjundiah, R.S.; Rao, S.A. On the modern deep learning approaches for precipitation downscaling. *Earth Sci. Inform.* **2023**, *1*–14. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.