

Article

Simulation and Reconstruction of Runoff in the High-Cold Mountains Area Based on Multiple Machine Learning Models

Shuyang Wang¹, Meiping Sun^{1,2,*}, Guoyu Wang¹, Xiaojun Yao^{1,2} , Meng Wang³ , Jiawei Li¹, Hongyu Duan¹ , Zhenyu Xie¹, Ruiyi Fan¹ and Yang Yang¹

¹ Department of Geography and Environment Sciences, Northwest Normal University, Lanzhou 730070, China; 202223072@nwnu.edu.cn (S.W.); 2020222779@nwnu.edu.cn (G.W.); xj_yao@nwnu.edu.cn (X.Y.); 2021222956@nwnu.edu.cn (J.L.); 2020120154@nwnu.edu.cn (H.D.); 202223106@nwnu.edu.cn (Z.X.); 2022212932@nwnu.edu.cn (R.F.); 2022212955@nwnu.edu.cn (Y.Y.)

² Key Laboratory of Resource Environment and Sustainable Development of Oasis, Lanzhou 730070, China

³ Chemistry and Chemical Engineering, Chongqing University of Science and Technology, Chongqing 401331, China; wangmeng@cqust.edu.cn

* Correspondence: sunmeiping1982@nwnu.edu.cn; Tel.: +86-188-9370-2035

Abstract: Runoff from the high-cold mountains area (HCMA) is the most important water resource in the arid zone, and its accurate forecasting is key to the scientific management of water resources downstream of the basin. Constrained by the scarcity of meteorological and hydrological stations in the HCMA and the inconsistency of the observed time series, the simulation and reconstruction of mountain runoff have always been a focus of cold region hydrological research. Based on the runoff observations of the Yurungkash and Kalakash Rivers, the upstream tributaries of the Hotan River on the northern slope of the Kunlun Mountains at different time periods, and the meteorological and atmospheric circulation indices, we used feature analysis and machine learning methods to select the input elements, train, simulate, and select the preferences of the machine learning models of the runoffs of the two watersheds, and reconstruct the missing time series runoff of the Kalakash River. The results show the following. (1) Air temperature is the most important driver of runoff variability in mountainous areas upstream of the Hotan River, and had the strongest performance in terms of the Pearson correlation coefficient (ρ_{XY}) and random forest feature importance (FI) ($\rho_{XY} = 0.63$, FI = 0.723), followed by soil temperature ($\rho_{XY} = 0.63$, FI = 0.043), precipitation, hours of sunshine, wind speed, relative humidity, and atmospheric circulation were weakly correlated. A total of 12 elements were selected as the machine learning input data. (2) Comparing the results of the Yurungkash River runoff simulated by eight machine learning methods, we found that the gradient boosting and random forest methods performed best, followed by the AdaBoost and Bagging methods, with Nash–Sutcliffe efficiency coefficients (NSE) of 0.84, 0.82, 0.78, and 0.78, while the support vector regression (NSE = 0.68), ridge (NSE = 0.53), K-nearest neighbor (NSE = 0.56), and linear regression (NSE = 0.51) were simulated poorly. (3) The application of four machine learning methods, gradient boosting, random forest, AdaBoost, and bagging, to simulate the runoff of the Kalakash River for 1978–1998 was generally outstanding, with the NSE exceeding 0.75, and the results of reconstructing the runoff data for the missing period (1999–2019) could well reflect the characteristics of the intra-annual and inter-annual changes in runoff.

Keywords: feature analysis; Hotan River Basin; high-cold mountains area; machine learning; runoff simulation and reconstruction



Citation: Wang, S.; Sun, M.; Wang, G.; Yao, X.; Wang, M.; Li, J.; Duan, H.; Xie, Z.; Fan, R.; Yang, Y. Simulation and Reconstruction of Runoff in the High-Cold Mountains Area Based on Multiple Machine Learning Models. *Water* **2023**, *15*, 3222. <https://doi.org/10.3390/w15183222>

Academic Editors: Yaning Chen, Zhi Li, Weili Duan, Chenggang Zhu, Gonghuan Fang and Yupeng Li

Received: 6 August 2023

Revised: 28 August 2023

Accepted: 31 August 2023

Published: 10 September 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

Global warming and increased human activity have led to changes in water resources, with significant ecological, environmental, and socio-economic impacts that are of widespread concern to the international community [1]. Glaciers are abundant in the high-cold mountains area (HCMA), which are the main source of freshwater resources in

the arid and semi-arid downstream areas [2,3]. Glaciers in many of the world's HCMAs have been observed to be melting significantly, and river runoff has increased such as the Kumalak River and Tuoshikan River, typical watersheds in the Tianshan Mountains of China, where runoff has increased by $1.5 \times 10^8 \text{ m}^3$ and $3.3 \times 10^8 \text{ m}^3$, respectively, over the past 50 years, leading to greater flooding [4–6]. Arid and semi-arid mountainous areas in northwest China are highly vulnerable to glacial snowmelt flooding under extreme climate change [7]. Floods can cause damage to local transportation facilities, downstream river runoff processes, and related hydrological forecasting systems such as the bursting of the Kayagil glacial dam in the Kunlun Mountains of Xinjiang [8,9]. Achieving sustainable use and scientific management of water resources in the HCMA is directly linked to the green and high quality development of the region downstream [10]. Runoff simulation and reconstruction can provide a rational decision-making basis for optimizing the allocation and use of water resources in the HCMA, thus reducing and preventing the occurrence of natural disasters such as floods, which is of great importance for ensuring downstream ecological security and promoting economic and social development [11,12].

To perform runoff simulation and reconstruction, commonly used methods include the Mann–Kendall trend test [13], wavelet reconstruction analysis [14], Pettitt mutation test [15], empirical orthogonal function analysis [16], bidirectional limit learning machines [17], and the traditional physical models SWAT [18] and Liuxihe [19], among others. Unfortunately, hydrological modeling is relatively difficult in HCMA regions of China where only sparse hydrometeorological data are available [20]. The spatial heterogeneity of the basin and the variable hydrological characteristics and morphology of the HCMA result in runoff time series that are nonlinear and nonstationary, making it difficult to accurately predict runoff and capture the characteristics of runoff changes [21,22]. The above methods require parameterization with sufficient data to achieve the best prediction and simulation results, and are not robust and cumbersome, making it difficult to provide scientifically accurate advice [23,24]. Therefore, there is an urgent need to explore a methodology that can accurately predict runoff in data scarce areas in a climate change environment. ML excels at dealing with nonlinear and nonsmooth data and has a strong ability to generalize, learn, and process high-dimensional data, which is widely used in hydrological forecasting [25,26]. For example, Rizeei et al. [27] used a random forest model to predict the runoff preventive flooding in the Damansara catchment in Malaysia; Langhammer [28] modeled different types of flood runoff in the upper Vydra Basin based on support vector machines (SVM) combined with hydrometeorological wireless sensor networks. Typical methods include random forest [29], gradient boosting [30], support vector machines [31], among others. Random forests independently and randomly select a subset of samples to construct multiple decision trees for prediction that can handle high-dimensional, nonlinear regression problems [32]. Gradient boosting uses the iterative principle to integrate a set of poor learners to learn and predict, dealing with nonlinear data with high interpretability and accuracy [33]. Support vector machines are based on statistical theory and the principle of minimum structural risk, and can solve problems such as nonlinearity and indivisibility in low-dimensional spaces [34]. In contrast to traditional models, machine learning models form a memory language based on their own black-box model training data, without considering the complex and variable physical processes in the catchment, which significantly improves the medium and long-term prediction performance of high-latitude nonlinear time series. However, there is no absolutely optimal machine learning model for the simulation of runoff, and it is necessary to predict runoff using multiple models such as random forest and gradient boosting and compare their performance. The Hotan River is one of the three remaining sources of the Tarim River, a representative inland river in the arid and semi-arid regions of northwest China [35]. The Yurungkash and Kalakash Rivers are tributaries of the Hotan River Basin, and their runoff is an important factor in providing water for downstream industry and agriculture, maintaining ecological balance, and limiting the extent of economic development in the region [36,37]. The scarcity of hydrological and meteorological stations in the HCMA makes it difficult to accurately predict runoff [38].

Runoff modeling can optimize the management of water resources in the Hotan River Basin and reduce losses due to droughts and floods. The third Xinjiang scientific team, tasked with exploring scientific models and methods of sustainable development of large inland river basins, carried out an assessment of the ecological and environmental quality of the Tarim River Basin. However, they found that the amount of water increased by glacial ablation on the north slope of the Kunlun Mountains was insufficient to explain the reasons for the significant increase in lakes in the basin, and that there were some shortcomings in the research on the mechanism of water area change and runoff change in mountainous areas. For this purpose, we used the Pearson correlation coefficient and random forest feature importance ranking methods to explore the correlation and degree of influence of different environmental factors on runoff in the Hotan River Basin on the northern slope of the Kunlun Mountains. Based on models from eight typical machine learning regression domains, random forest, gradient boosting, AdaBoost, bagging, support vector regression, ridge, K-nearest neighbor, and linear regression, and combining long time series data of temperature, precipitation, humidity, sunshine hours, and typical atmospheric circulation, the runoff simulation and reconstruction were performed, and the aim was to make up for the lack of data and to explore trends in runoff. The main objectives of this study were as follows. (1) Accomplish preferences for input parameters such as air temperature and precipitation in the Hotan River Basin and determine the influence of environmental factors on runoff in the basin. (2) Realization of the runoff simulation of the Yurungkash River for 1999–2019 and the evaluation of the model quality and simulation accuracy. (3) Simulation of the runoff of the Kalakash River for the years 1978–1998 and reconstruction of the runoff for the years 1999–2019 for the years with missing data were implemented, and the results of the simulation and reconstruction were analyzed. This study will provide methodological references to bridge the gaps of the Third Xinjiang Scientific Expedition and realize water resource management in arid and semi-arid regions.

2. Research Area and Data

2.1. Research Area

In this study area, the Yurungkash and Kalakash Rivers are two tributaries in the middle and upper reaches of the Hotan River Basin (Figure 1). The Yurungkash River is geographically located between $81^{\circ}41' \sim 79^{\circ}22'$ E longitude and $38^{\circ}15' \sim 35^{\circ}25'$ N latitude, originating from the northern foothills of the Kunlun Mountains, the total length is 513 km, the area is $1.98 \times 10^4 \text{ km}^2$, the average air temperature of the basin is 10.6°C , the precipitation is 38.4 mm, and the average runoff of the Tongguziluoke Hydrological Station is $21.95 \times 10^8 \text{ m}^3$. There are more than 1300 glaciers in the Yurungkash River Basin, with a total area of 2958.31 km^2 , total coverage of 20.30%, and the glacier reserves are 410.32 km^3 [39]. The geographical location of the Kalakash River is between $77^{\circ}25' \sim 80^{\circ}$ E and $34^{\circ}52' \sim 38^{\circ}04'$ N, originating from the Tuanjie Peak in the Karakoram Mountains, the total length of 808 km, the source of the highest mountain elevation of 6662 m, the average temperature for many years 11.3°C , the annual precipitation of 36.5 mm, Wuluwati Hydrological Station cross-section of the measured multi-year average runoff of $21.51 \times 10^8 \text{ m}^3$. For the Kalakash River Basin, there are more than 1900 glaciers with a total area of 2163.17 km^2 , total coverage of 10.83%, glacier reserves 156.09 km^3 (Table 1) [40]. The two tributaries pass through the oasis plains of Hotan, Karakax and Lop Counties, connect with the Hotan River at Kuoshilashi, then flow for 319 km and join the Tarim River at Xiaojiake [41]. The simulation and reconstruction of runoff in the Hotan River Basin can provide recommendations for future water resources planning and management in the mountainous basin, which is of great significance for safeguarding the ecological balance of the green corridor zone of the Hotan River as well as promoting the socio-economic development of arid and semi-arid regions in China [42,43].

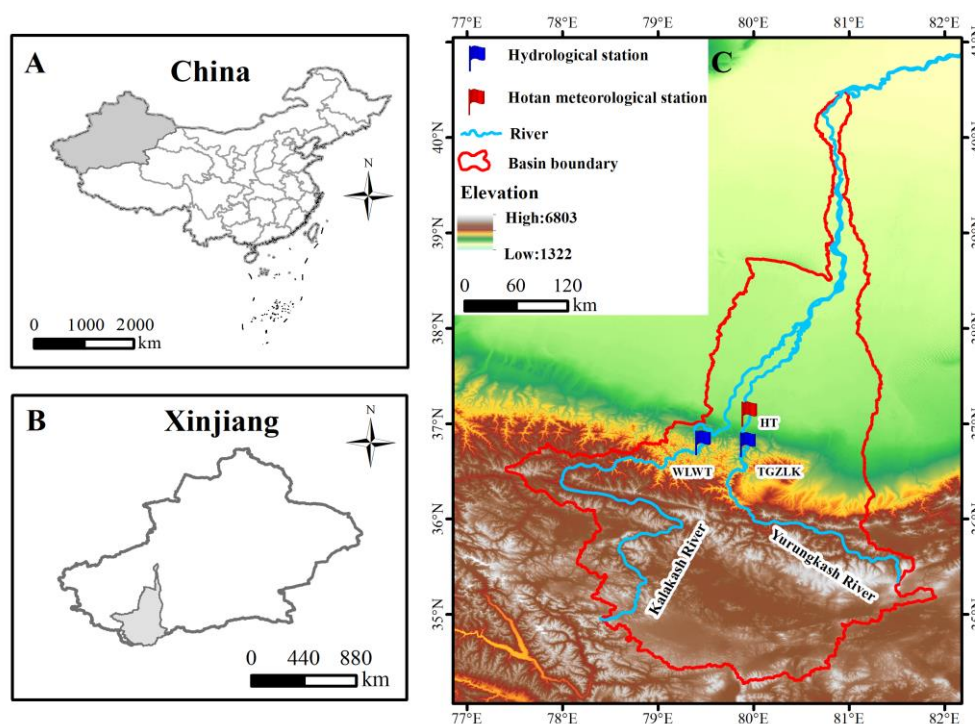


Figure 1. The locations of the Xinjiang Uygur Autonomous Region (A) and the Hotan River Basin (B) in China are shown. The Hotan River Basin (C) includes the Hotan (HT) Meteorological Station, Tongguzilouke (TGZLK), and Wuluwati (WLWT) Hydrological Stations.

Table 1. Parameters of the study area.

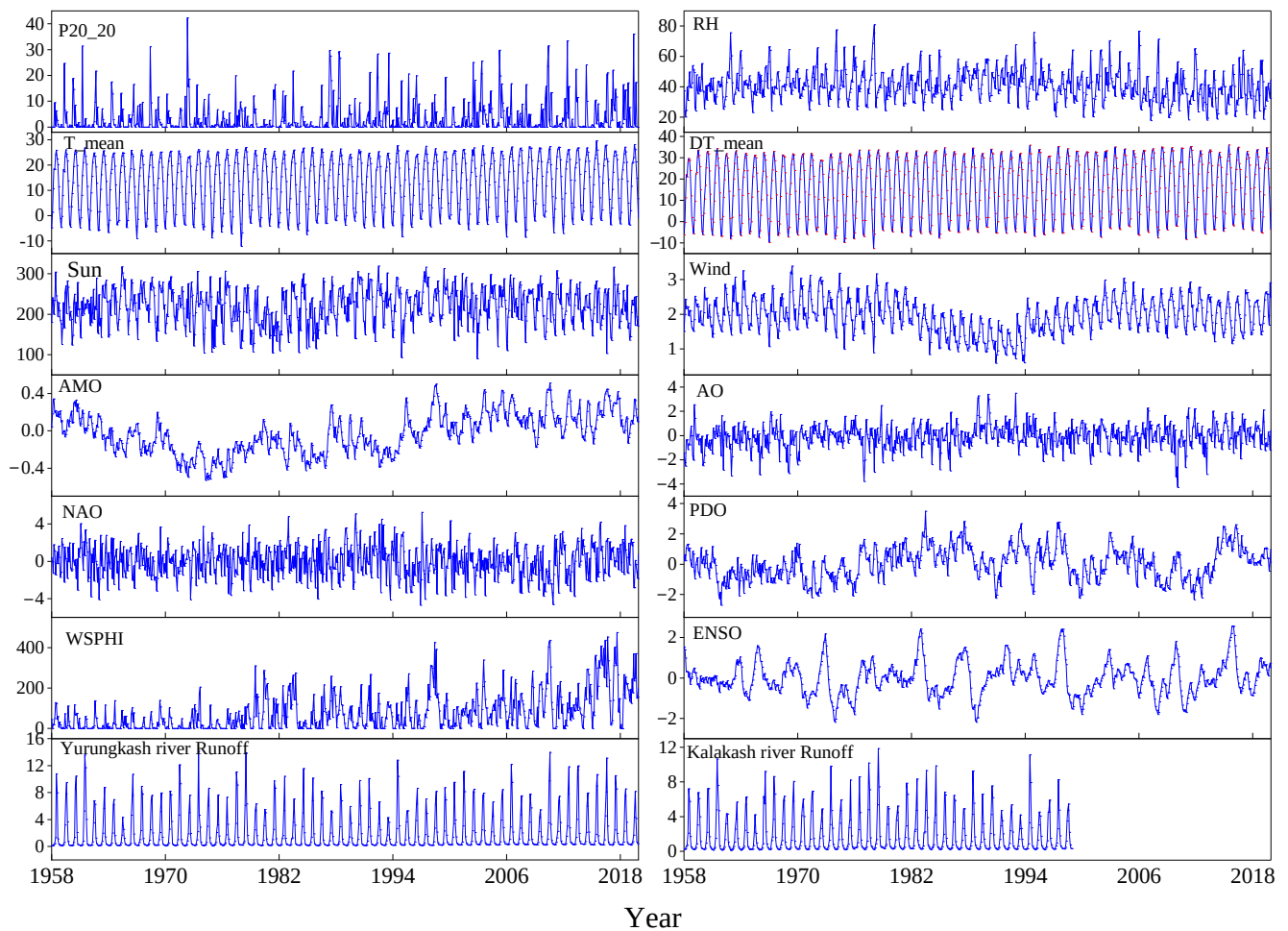
River Basin	Lengths (km)	Area ($\times 10^4$ km ²)	Temperature (°C)	Precipitation (mm)	Runoff ($\times 10^8$ m ³)	Glacier Area (km ²)
Yurungkash	513	1.98	10.6	38.4	21.95	2958.31
Kalakash	808	2.66	11.3	36.5	21.51	2163.17

2.2. Data

Important drivers of runoff change are temperature and precipitation [44,45]. Humidity and sunshine hours affect the effects of evapotranspiration, and wind speed accelerates glacial snowmelt, leading to changes in runoff [46]. Atmospheric circulation is linked to global climate change [47]. The input data used in this study were as follows (Table 2): the monthly meteorological observations were obtained from the Hotan Meteorological Station of the Xinjiang Uygur Autonomous Region of the National Center for Meteorological Data and Science (NCDC) of China, hydrological data were obtained from the Tongguzilok Hydrological Station (TGZLK) and the Wuluwati Hydrological Station (WLWT), and atmospheric circulation data were downloaded from the Global Climate Observing System (GCOS) (https://www.psl.noaa.gov/gcos_wgsp/, (accessed on 5 July 2023)). Specifically, it includes the monthly temperature (average air temperature, average soil temperature), total precipitation, average relative humidity, total sunshine hours, and average wind speed from 1958 to 2019; monthly runoff data of the Yurungkash River from 1958 to 2019; monthly runoff data of the Kalakash River from 1958 to 1998 (the runoff data from 1999 to 2019 are missing); monthly runoff data of the Yurungkash River from 1958 to 2019 El Niño-Southern Oscillation (ENSO), Pacific Decadal Oscillation (PDO), Arctic Oscillation (AO), Atlantic Multidecadal Oscillation (AMO), North Atlantic Oscillation (NAO), and Western Pacific Subtropical High Pressure Intensity (WPSHI) on a monthly scale. A time series plot of the dataset is shown in Figure 2.

Table 2. Type of data.

Data Types	Input Data Name	Time Span	Obtaining Sources
Meteorological data	Air temperature (T_mean)	1958–2019	Hotan Meteorological Station (HT)
	Soil temperature (DT_mean)		
	Total precipitation (P20_20)		
	Relative humidity (RH)		
	Sunshine hours (Sun)		
Hydrological data	Yurungkash River runoff	1958–2019	Tongguziluoke Hydrographic Station (TGZLK)
	Kalakash River runoff	1958–1998	Wuluwati Hydrographic Station (WLWT)
Atmospheric circulation data	El Niño-Southern Oscillation (ENSO)	1958–2019	Global Climate Observing System (GCOS) https://www.psl.noaa.gov/gcos_wgsp/ (accessed on 5 July 2023)
	Pacific Decadal Oscillation (PDO)		
	Arctic Oscillation (AO)		
	Atlantic Multidecadal Oscillation (AMO)		
	North Atlantic Oscillation (NAO)		
	Western Pacific Subtropical High Pressure Intensity (WPSHI)		

**Figure 2.** Time series of the datasets.

3. Research Methods

The flow of the study is shown in Figure 3. Firstly, the basic variables are selected as the model input set including the general circulation index, meteorological data, and runoff data; secondly, the model construction is carried out where the input parameters of the model are used to perform feature selection using the Pearson correlation coefficient and random forest feature importance ranking to judge the influence of the parameters on the model simulation, in the meantime, the input data are normalized to transform nonstationary data into stationary series and remove outliers, where 70% of the input data are divided as the training data and 30% as the simulation data, and then predicted based on eight ML models; finally, the runoff simulation and reconstruction results are obtained and evaluated (NSE, PBIAS, RMSE, MAE). In the subsequent subsections, we describe each step in more detail.

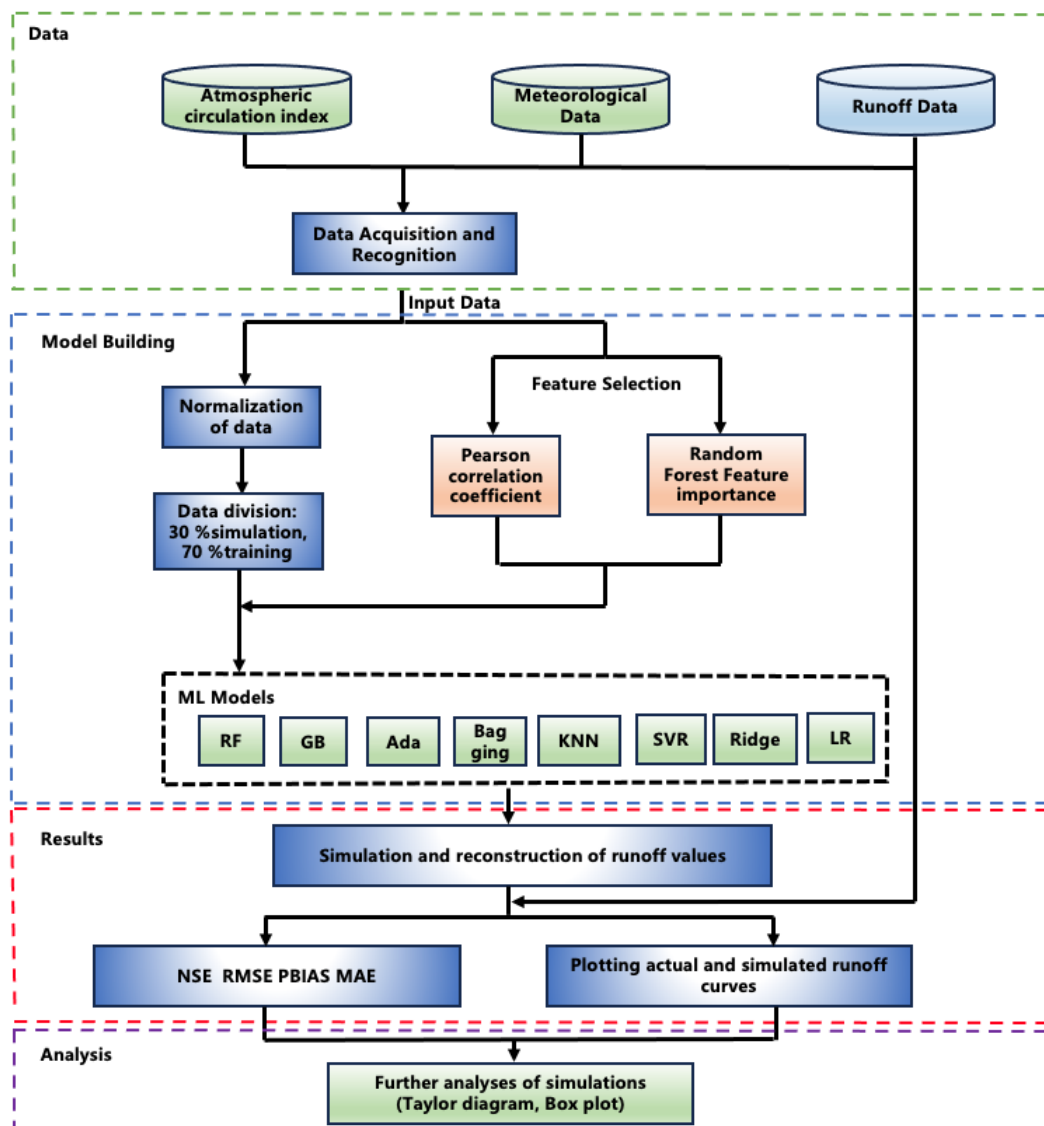


Figure 3. Flowchart of the machine learning models to simulate runoff.

3.1. Runoff Simulation and Reconstruction Modelling

We used eight machine learning (ML) regression models for runoff simulation and reconstruction including random forest, gradient boost, support vector regression, AdaBoost, KNN, bagging, ridge, and linear regression. The application relies on Python 3.9 and the “Scikit-Learn” package. The methods are described in Table 3.

Table 3. Principles of eight machine learning (ML) methods and their strengths and weaknesses.

ML Models	The Core Idea	Strengths and Weaknesses	Reference
Random Forest (RF)	Randomly and independently select a subset of samples to construct multiple decision trees for training, input unknown data, predict each decision tree, and use voting or averaging to obtain the final regression results.	It can better prevent the overfitting phenomenon and overcome the problem of too large a feature dimension, with simple model structure, short training time, high efficiency, strong generalization ability, and good robustness. However, for the sample set with too much data noise, it is easy to produce the overfitting phenomenon.	[48–50]
Gradient Boosting	The training process first finds a model with weak prediction accuracy, gradually reduces the residuals by adding a predictor, and calculates the residual value between the predicted and actual values of the model to achieve the purpose of improving accuracy, using the iterative principle to use the appropriate loss function, develops a strong learner based on the weak learner, and performs prediction simulation.	The training effect is better, not easy to produce overfitting, with the advantages of high interpretability, high learning efficiency, minimal prediction error, and high stability. However, it requires careful parameter tuning and longer training time.	[51,52]
Support Vector Regression (SVR)	Using support vector machines to fit curves for regression analysis, finding a plane to which all the data in the set are closest, minimizing the risk to the expected value, is a machine learning regression algorithm based on support vector machines.	It can effectively solve the regression problem of high-dimensional features, only needing to use part of the support vector to do the decision of the hyperplane, with high accuracy and resolution. However, it is very sensitive to missing data and not very applicable when the sample size is very large.	[53,54]
AdaBoost	Multiple weakly learned classifiers are learned by changing the weights of the training samples, and then these weakly learned classifiers are assembled to form a strongest learner for linear fitting to achieve the purpose of predictive simulation.	It can solve multi-class single-label and multi-label problems with high accuracy, highly flexible in use, and fully considers the weight of each classifier. However, the number of classifiers is not well set, and the imbalance of experimental data will lead to a decrease in prediction accuracy and a longer training time.	[55,56]
K-Nearest Neighbor (KNN)	KNN scans the set of training samples to find the training sample that is most like the test sample, and then votes to determine the class of the test sample based on the class of the most similar training sample, or votes weighted by the degree of similarity between each sample and the test sample to obtain the result.	KNN is the simplest model in the learner. Based on the KNN regression algorithm, there is no need to consider boundary instances. However, using KNN is more computationally intensive, has low prediction accuracy when the samples are unbalanced, slow to predict, and not very interpretable.	[57–59]
Ridge	An improvement on the method of least squares estimation. The core idea is to determine the value of the regular term coefficient parameter K. It is dedicated to solving the covariance data partitioning problem and is a regularized regression model.	Enough to effectively reduce the data overfitting phenomenon, the prediction of unknown data is more robust and the obtained regression coefficients are more in line with the mathematical reality, but its fitting ability is easily limited and the underfitting phenomenon may occur.	[60,61]

Table 3. Cont.

ML Models	The Core Idea	Strengths and Weaknesses	Reference
Bagging	The input randomized uniformly selected dataset is trained in multiple rounds to construct weak learners with differences and parallel relationships, which are combined to obtain the final strong learner.	Bagging can be used directly to solve multi-classification and regression problems; by reducing the variance of the classifier, it improves the flourish error and can effectively prevent overfitting. However, underfitting can occur.	[62,63]
Linear Regression	Based on regression analysis in mathematics, a straight line is used to describe the relationship more accurately between one or more independent variables and the dependent data. The input data are trained and processed in an algorithmic language to produce a simple prediction value.	The algorithm is simple, fast, and interpretable; however, it can only be used for regression problems, lacks some logic, has a low accuracy of predicted value, and is prone to overfitting.	[64]

3.2. Feature Selection

The selection of input features affects the accuracy of the results of machine learning (ML) model training. Feature selection is the process of selecting a subset of features with strong discriminative power according to a given criterion and feature in each feature dataset, which is critical to the accuracy of the model training results. By selecting from all the features those that are beneficial to the machine learning algorithm, and determining which influence of the features is valid or unknown, the training efficiency of the machine learning model can be improved [65]. In this study, we determined the effect of feature parameters on runoff based on the Pearson correlation coefficient and random forest feature importance ranking method to make an excellent selection of input features. The principle is as follows.

The Pearson correlation coefficient is applicable to the correlation analysis of continuous variables, it can characterize the degree of linear correlation between the data, the range of values is $[-1, 1]$, the value greater than 0 indicates that the two vectors are positively correlated, less than 0 is negatively correlated, the value equal to 0, then the two vectors are not correlated, and the closer the absolute value is to 1, the stronger the correlation [66]. The specific formula is as follows:

$$\rho_{XY} = \frac{Cov(X, Y)}{\sqrt{D(X)} \times \sqrt{D(Y)}} = \frac{E(X - EX) \times E(Y - EY)}{\sqrt{D(X)} \times \sqrt{D(Y)}} \quad (1)$$

where X and Y refer to two independent characteristics, E is the mathematical expectation, D is the variance, $\sqrt{D(X)}$, $\sqrt{D(Y)}$ is the standard deviation, $Cov(X, Y)$ is the covariance of the sum of the random variables, and ρ_{XY} is the correlation coefficient of X and Y .

The key idea of random forest feature importance selection is to transform feature parameters into random numbers, calculate their impact on model accuracy, and measure the importance of the feature parameter based on the average reduced accuracy value obtained from multiple calculations, the degree of contribution of quantitative descriptive features to classification or regression, where the higher the value, the higher the importance of the corresponding features [67]. The selected features are used to make a decision in the internal node, and it divides the dataset into two separate sets with similar responses [68]. The calculation starts by using the corresponding out-of-bag (OOB) data for each decision tree in the random forest to calculate the out-of-bag error, detected Err_{OOB1} , then randomly adds noise interference to the features of all the samples of the out-of-bag data OOB,

computes its out-of-bag data again, detected $ErrOOB2$. Suppose there are N decision trees in the random forest, the formula is as follows [69]:

$$FI \text{ (Feature importance)} = \frac{\sum (ErrOOB2 - ErrOOB1)}{N} \quad (2)$$

3.3. Evaluation Parameters

Evaluating model performance is a key step in judging the success or failure of runoff simulations [70]. The Nash–Sutcliffe efficiency coefficients (NSE) are calculated from the time series of the runoff simulation results and real observations of runoff (Equations (3)), and are commonly used to evaluate the quality of hydrological models. The value of NSE ranges from negative infinity to 1. The closer it is to 1, the closer the model simulation results are to the true value and the better the model quality [71]. The NSE mathematical formula is as follows [72]:

$$NSE = 1 - \frac{\sum_{t=1}^T (Q_0^t - Q_m^t)^2}{\sum_{t=1}^T (Q_0^t - \bar{Q}_0)^2} \quad (3)$$

In addition to assessing the model quality using the Nash coefficients, a quantitative assessment of model performance is required by assessing the average tendency of the model simulated data to be greater or less than the actual observed data using the percentage bias ($PBIAS$). The optimum value of $PBIAS$ is 0, and the closer to 0, the smaller the deviation of the model results. A hydrological runoff simulation with a $PBIAS$ less than $\pm 10\%$ is excellent, the results in the range of $\pm 10\sim\pm 15\%$ are well, and the model results are reliable in the range of less than $\pm 25\%$ [73,74]. $PBIAS$ mathematical formulae [75]:

$$PBIAS = \left[\frac{\sum_{t=1}^T (Q_0^t - Q_m^t)}{\sum_{t=1}^T Q_0^t} \right] \times 100 \quad (4)$$

Root mean square error ($RMSE$) is used to reflect the overall deviation between the model predictions and actual observations; the smaller the $RMSE$ value, the more accurate and precise the model simulation [76]. The $RMSE$ mathematical formula is:

$$RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^T (Q_0^t - Q_m^t)^2} \quad (5)$$

Mean absolute error (MAE) is used to describe the average degree of deviation of the model's simulated predictions from the actual observed values, which can accurately reflect the true magnitude of the prediction error. The smaller the MAE , the better the model prediction, and the closer the result is to the real value [77]. Equation (6) is used to calculate the MAE :

$$MAE = \frac{1}{T} \sum_{t=1}^T (Q_0^t - Q_m^t) \quad (6)$$

In Equations (3)–(6), Q_0 refers to the actual observed runoff, Q_m refers to the runoff simulated by the model, $\bar{Q}_0$ represents the mean of the actual observed value, the t superscript represents the moment t , and T represents the total number of time series.

4. Results and Analyses

4.1. Feature Analysis

Patro et al. [78] showed that wind speed and humidity can accelerate glacial snow melt. Taylor et al. [79] showed that atmospheric circulation effects have an impact on global air temperature and precipitation. The air temperature and precipitation are the main factors directly responsible for changes in runoff in most rivers around the globe [80]. Therefore, we considered introducing parameters such as meteorological factors (air temperature (T_{mean}), soil temperature (DT_{mean}), precipitation ($P20_{20}$), wind speed ($Wind$), relative

humidity (RH), sunshine hours (Sunshine)) and atmospheric circulation effects (ENSO, AO, AMO, NAO, PDO, WPSHI) in the Hotan River Basin for characterization and optimization.

4.1.1. Pearson Correlation Coefficient

The Pearson correlation coefficients between the two variables were calculated (Figure 4). We found some correlations between some variables such as the correlation between the T_mean and RH, Sunshine, and Wind reached 0.48, 0.58, and 0.57, respectively; the correlation between RH and Wind reached 0.53; the correlation between WPSHI and AMO reached 0.49; the correlation between PDO and ENSO reached 0.42. There are two ways to solve the multiple covariate problem (i.e., direct exclusion of variables and increasing the sample size). We used the former method to simulate the runoff and found that the NSE decreased by 0.1–0.15, which further proved that the existence of multicollinearity between the variables did not affect the runoff simulation results of this study. Comparison of the Pearson correlations between different input features and runoff is shown in Figure 5a. For a more intuitive analysis, the Pearson correlation coefficient values were absolute. The meteorological factors DT_mean ($\rho_{XY} = 0.63$) and T_mean ($\rho_{XY} = 0.63$) were strongly correlated with runoff, P20_20 ($\rho_{XY} = 0.18$), Wind ($\rho_{XY} = 0.21$), and Sunshine ($\rho_{XY} = 0.22$) were moderately correlated, and RH ($\rho_{XY} = 0.02$) was weakly correlated. The Hotan River Basin belongs to the HCMA, with less clouds, more glaciers and snow, scarce precipitation, and runoff mainly from glaciers and snowmelt recharge [81]. Therefore, the T_mean and DT_mean are more important for runoff than precipitation and had the strongest correlation. Wind speed and sunshine hours influenced the temperature changes and indirectly contributed to glacier snowmelt, with a slightly higher correlation than P20_20. RH had less of an effect on runoff due to the dry and cold climate in the HCMA. The atmospheric circulation effects of WPSHI ($\rho_{XY} = 0.24$) and AMO ($\rho_{XY} = 0.13$) were nearly moderately correlated with runoff, and PDO ($\rho_{XY} = 0.06$), ENSO ($\rho_{XY} = 0.04$), NAO ($\rho_{XY} = 0.03$), and AO ($\rho_{XY} = 0.01$) were weakly correlated with runoff. Atmospheric circulation can influence global climate change and had a non-negligible influence on the formation and generation of runoff in the HCMA, but the correlation was small and can be used as an input feature to improve the credibility of runoff predictions.

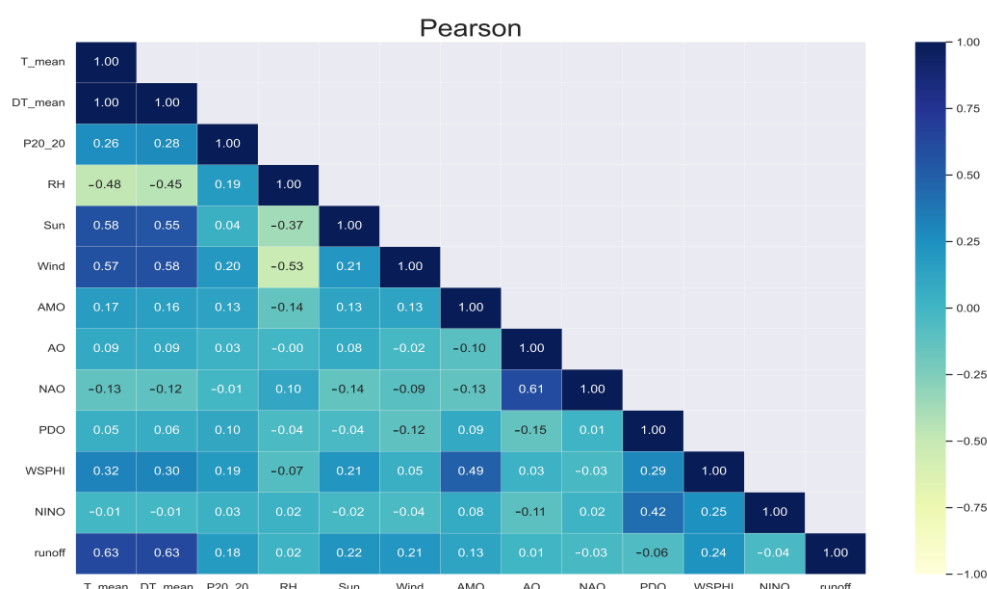


Figure 4. Pearson correlation coefficient matrix between two variables.

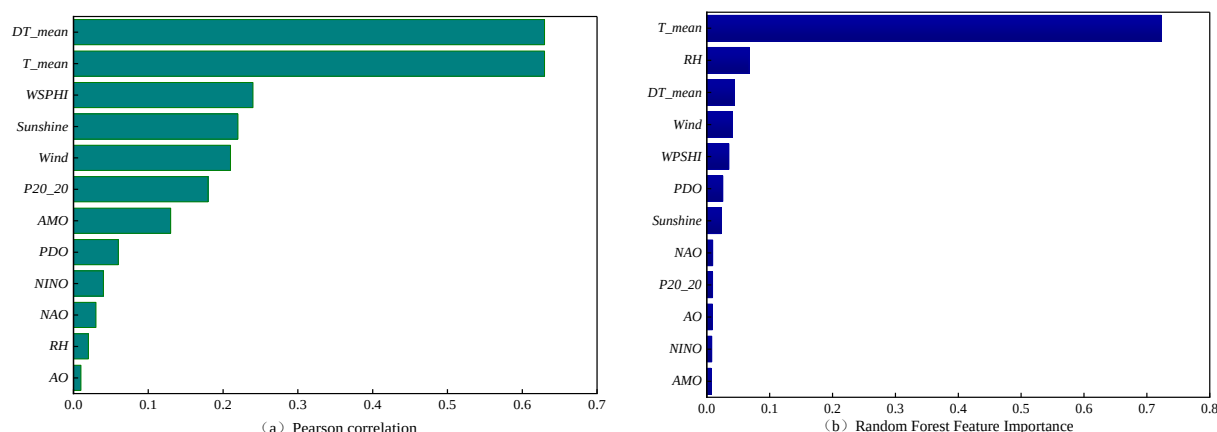


Figure 5. Feature analysis: Pearson correlation coefficient (a) and random forest feature importance (b).

4.1.2. Random Forest Feature Importance

Random forest feature importance ranking was used for 12 input parameters (Figure 4). The order was $T_mean > RH > DT_mean > Wind > WSPHI > PDO > Sunshine > NAO > P20_20 > AO > ENSO > AMO$. This shows that among the meteorological factors, T_mean ($FI = 0.723$) ranked much higher than DT_mean ($FI = 0.043$) with the strongest importance, RH ($FI = 0.067$) ranked second, $Wind$ ($FI = 0.041$) and $Sunshine$ ($FI = 0.023$) ranked moderately, and $P20_20$ ($FI = 0.009$) ranked low. In the atmospheric circulation index, the features of $WSPHI$ ($FI = 0.035$) and PDO ($FI = 0.025$) were moderately ranked, and the importance of the features of AMO ($FI = 0.007$), NAO ($FI = 0.009$), $ENSO$ ($FI = 0.008$), and AO ($FI = 0.009$) was weak. The results show that air temperature is still the most important parameter contributing to the runoff process simulated in the random forest and had a much greater influence on the model simulation results than surface temperature and precipitation. The atmospheric circulation factors ranked lower, but were still of some importance, which is also consistent with the results of the Pearson's correlation.

Both the Pearson correlation coefficient and the random forest feature importance ranking indicate that temperature is the most important parameter influencing runoff, which is closely related to the climatic conditions of the Hotan River Basin as a general trend of global warming. The Hotan River Basin is arid and has low rainfall, so precipitation is weakly correlated. $Wind$, RH , and $Sunshine$ showed a better correlation in the feature, which could increase the credibility of the model simulation results. Atmospheric circulation was weakly correlated with runoff and had a low order of influence on the runoff modeling performance. Twelve factors were selected as machine learning input data for the runoff prediction study including air temperature, precipitation, sunshine, and atmospheric circulation.

4.2. Runoff Simulation of the Yurungkash River

Eight ML regression methods including random forest, gradient boosting, support vector regression (SVR), AdaBoost, KNN, bagging, ridge, and linear regression were used to simulate the runoff from 1999 to 2019 using the monthly average meteorological data, atmospheric circulation data, and runoff data of the Yurungkash River from 1958 to 1999 as training samples, and were then combined with the actual observed runoff data to calculate the evaluation parameters and validate the results.

Time series curves of the runoff simulation and actual observation data were plotted (Figure 6). The predicted and actual time series curves between different machine learning methods had a certain fit, but there were differences in the degree of coincidence. The simulated runoff curves of the random forest, gradient boosting, AdaBoost, and bagging models had a high degree of fit with the measured curves, which could clearly reflect the intra-annual and inter-annual trends of the Yurungkash River runoff. The runoff peaks all occurred at the same time and the modeled values were slightly lower than the measured

values. Comparing the box plots of the simulated and measured runoff from different models (Figure 7A), it can be seen that the difference in box width between the simulated runoff from random forest (RF), gradient boosting (GB), bagging, and AdaBoost (Ada) and the measured runoff data (Real) was small. This indicates that the simulated data of the four models had a similar degree of fluctuation with the measured runoff, and the simulated values were close to the real values and the simulation quality was well done. The degree of fluctuation was similar to that of the measured runoff, and the simulated values were close to the real values, with good simulation quality. The simulated runoff values of KNN, SVR, ridge and linear regression (LR) were less than 0, and the end lines as well as the width of the box were longer, which deteriorated the simulation effect.

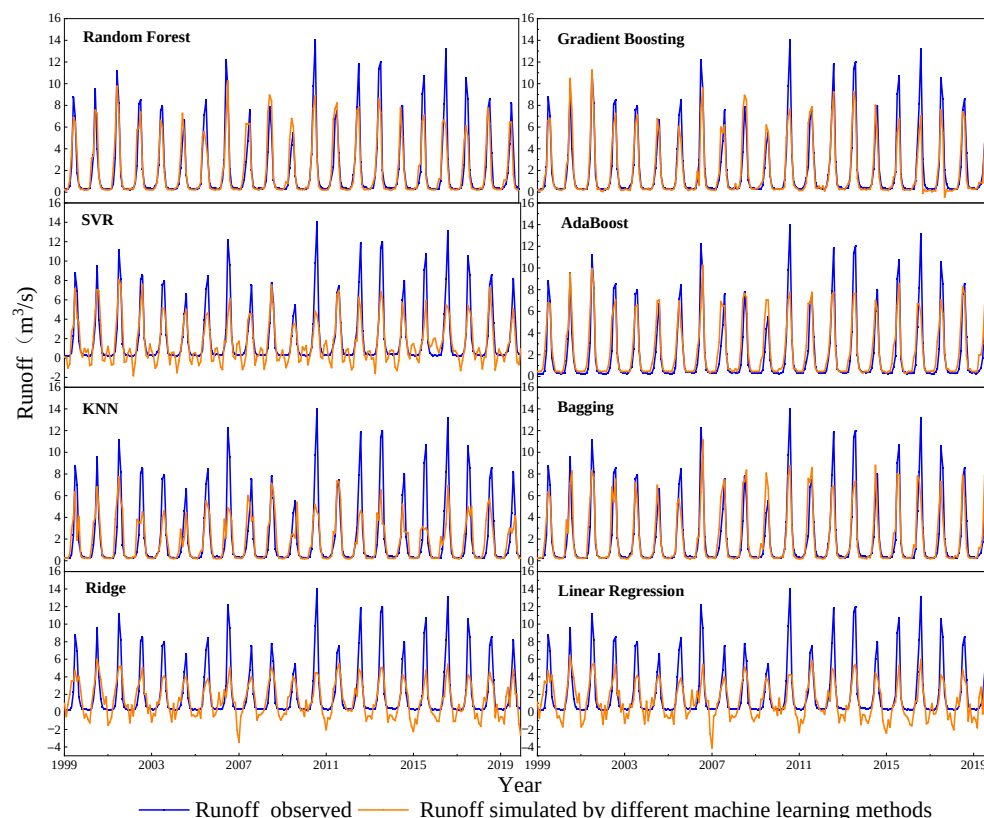


Figure 6. Comparison of the time series curves of the simulation results and actual observations of runoff in the Yurungkash River from 1999 to 2019.

Calculation of the model simulation evaluation parameters was as follows (Table 4). All eight methods showed Nash–Sutcliffe efficiency coefficients (NSE) between 0 and 1. The closest to 1 was gradient boosting (NSE = 0.84), followed by the random forest (NSE = 0.82), AdaBoost (NSE = 0.78), bagging (NSE = 0.78), ridge (NSE = 0.530), and linear regression (NSE = 0.51) models, which had low NSE and were not as credible as the other method. Random forest (PBIAS = 4.89%) and gradient boost (PBIAS = 9.44%) regression models with PBIAS less than $\pm 10\%$ performed well, random forest was the closest to 0 with the best trend in model simulation, bagging (PBIAS = 11.07%), AdaBoost (PBIAS = -14.42%), and KNN (PBIAS = 13.94%) performed well with a PBIAS between $\pm 15\%$, SVR (PBIAS = 24.95%), linear regression (PBIAS = 39.99%), and ridge (PBIAS = 34.13%) had larger PBIAS. The smaller the value of the RMSE and MAE, the closer the model prediction result is to the real value, so the higher the model accuracy. Gradient boosted regression (RMSE = 1.24, MAE = 0.65) had the smallest value, random forest (RMSE = 1.32, MAE = 0.70), AdaBoost (RMSE = 1.38, MAE = 0.81), and bagging (RMSE = 1.42, MAE = 0.75) had smaller errors and performed better with higher confidence. The maximum values of RMSE and MAE were seen for the linear regression model (RMSE = 2.14, MAE = 1.48), which increased the linear

regression RMSE and MAE values by 0.90 and 0.83, respectively, with a large difference compared to the gradient boosting method, and ridge (RMSE = 2.11 and MAE = 1.42) was like the linear regression with poor model performance. Plotting the Taylor diagram of the model simulation evaluation parameters (Figure 7B) showed that gradient boosting was the best performing model, random forest, AdaBoost, and bagging regression models showed excellent performance, and the four remaining simulated methods performed poorly.

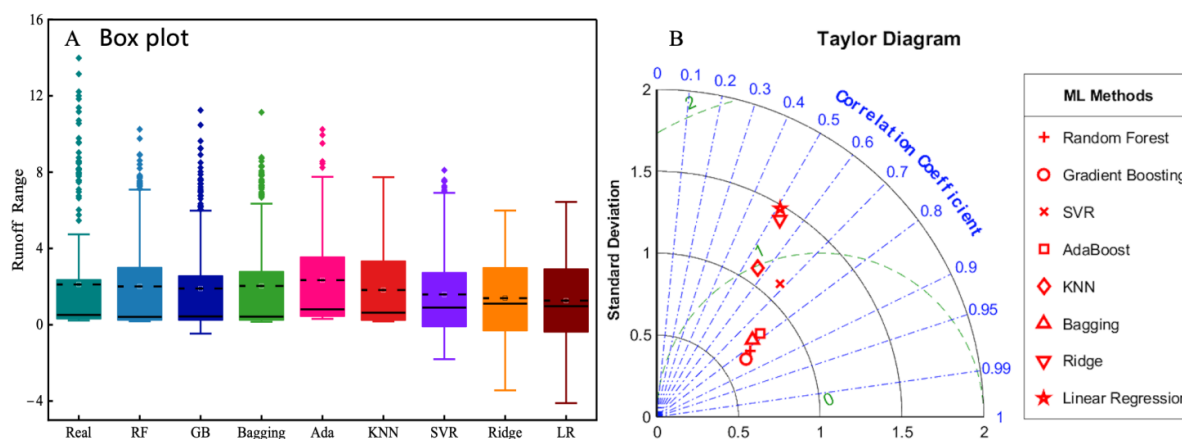


Figure 7. Box plot (A) of the eight ML models simulating runoff from the Yurungkash River versus the measured runoff and the Taylor diagram (B) of the evaluation parameters (NSE, RMSE, MAE) for the eight ML models.

Table 4. Accuracy of the simulated runoff by ML in the Yurungkash River.

ML Methods	NSE	PBIAS (%)	RMSE	MAE
Random Forest	0.82	4.89	1.32	0.70
Gradient Boosting	0.84	9.44	1.24	0.65
SVR	0.68	24.95	1.74	1.11
AdaBoost	0.78	−14.42	1.38	0.81
KNN	0.56	13.94	2.03	1.10
Bagging	0.78	11.07	1.42	0.75
Ridge	0.53	34.13	2.11	1.42
Linear Regression	0.51	39.99	2.14	1.48

ML regression methods could effectively simulate the river runoff of the Yurungkash River from 1999 to 2019, with differences in accuracy and quality between the different models. According to the runoff simulation curve and evaluation parameter calculation results synthesis, it was found that gradient boosting was the best performing model, random forest performed second best, and then AdaBoost and the bagging model; these four models for runoff simulation had high quality, precision, accuracy. The NSE of SVR was lower than the bagging regression model, the prediction was underestimated, and the performance was average. The NSE of KNN, ridge, and linear regression was poor, the prediction curve fit was minor, the percentage deviation was huge, the model prediction value was overall credible, but for the other several methods, the model quality was poor.

4.3. Runoff Simulation and Reconstruction of the Kalakash River

The simulation results of the Yurungkash River runoff found that the gradient boosting, random forest, AdaBoost, and bagging regression model simulation quality and accuracy were higher, and the simulation results were more accurate. Therefore, the above four methods were used to simulate and reconstruct the missing data of the Kalakash River runoff. For the Kalakash River, only the monthly mean meteorological data from 1958 to 2019 and the atmospheric circulation data and the runoff data from 1958 to 1999 were available. The training sample sequence was 1958–1977 and the simulation validation data

sequence was 1978–1998 in the runoff simulation process. The training sample sequence was 1958–1998 and the reconstruction sequence was 1999–2019 for the runoff reconstruction.

4.3.1. Runoff Simulation of the Kalakash River

Time series curves of the simulated and actual observed data of runoff in the Kalakash River were plotted (Figure 8). Compared to the training data of the Yurungkash River runoff simulation, the Kalakash River runoff simulation data decreased, and the PBIAS were all less than 0 but did not affect the accuracy of the simulation results. The simulated runoff curves and measured runoff curves from the four methods were in good agreement, the capture of the peak was slightly higher compared to the measured runoff, but the peaks were all at the same moment. Comparing the box plot of simulated runoff with measured runoff for the four models (Figure 9A), it was found that the simulated runoff such as gradient boosting (GB) had a shorter endline with the data of measured runoff (Real), the mean and median positions were similar to the values of the measured runoff, and it effectively reduced the number of anomalies, which provides a better quality of simulation.

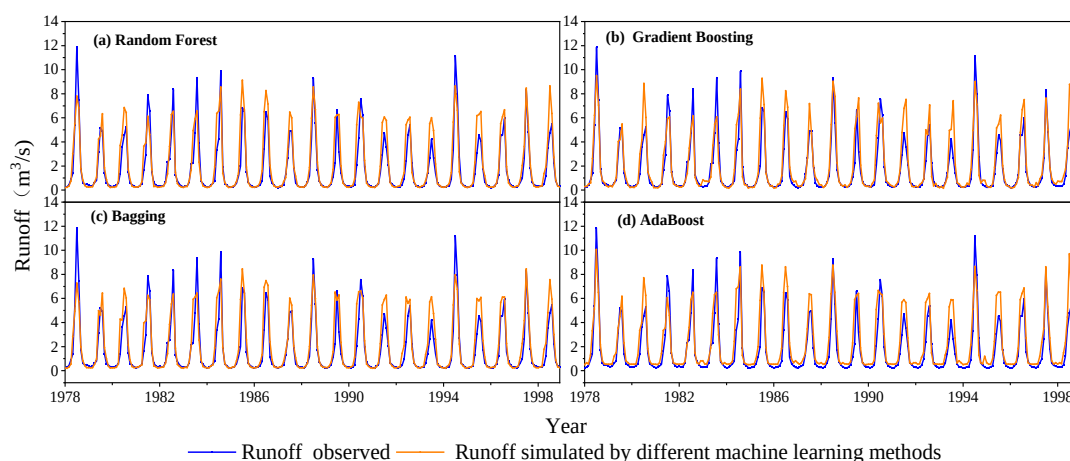


Figure 8. Comparison of the time series curves of the simulation results and actual observations of runoff in the Kalakash River from 1978 to 1998.

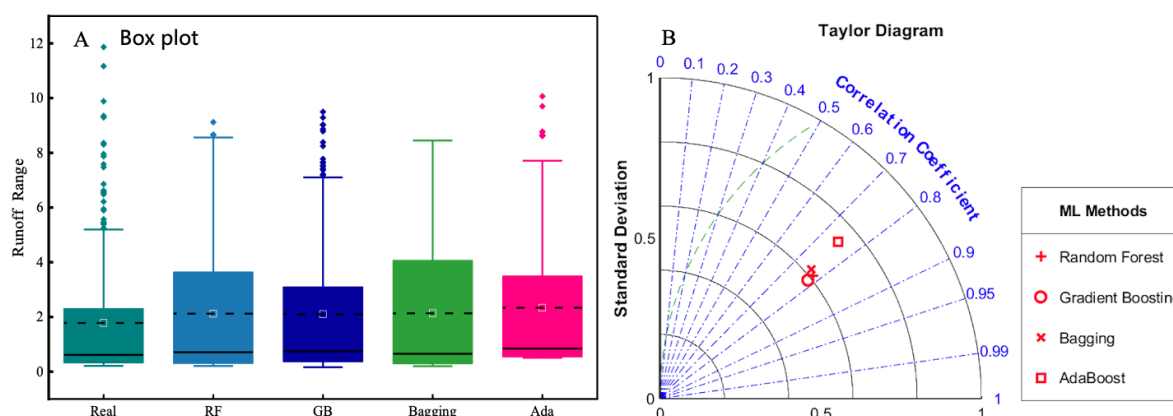


Figure 9. Box plot (A) of the four ML models simulating runoff from the Kalakash River versus the measured runoff and the Taylor diagram (B) of the evaluation parameters (NSE, RMSE, MAE) for the four ML models.

Analyzing the values of the NSE, RMSE, and MAE evaluation parameters (Table 5), it was found that random forest (NSE = 0.78, RMSE = 1.08, MAE = 0.61), gradient boosting (NSE = 0.78, RMSE = 1.06, MAE = 0.59), and bagging (NSE = 0.76, RMSE = 1.11, MAE = 0.62), AdaBoost (NSE = 0.75, RMSE = 1.13, MAE = 0.74)—four models used to simulate the Kalakash River runoff—also had high quality and accuracy, the model results

of the RMSE and MAE were smaller, the NSE was close to 1, and all were greater than 0.75. Plotting the Taylor diagram of the model simulation evaluation parameters (Figure 9B) showed that gradient boosting was the best performing model, random forest and bagging regression model performed well, and the AdaBoost model performed slightly lower than the other four methods. The results show that the runoff simulation of the Kalakash River based on the random forest, gradient boosting, bagging, and AdaBoost ML regression methods had high accuracy and superiority in the case of reduced training data, the time series curves of the real and simulated values showed excellent fit with each other, and the simulation results were credible.

Table 5. Accuracy of the simulated runoff by ML in the Kalakash River.

ML Methods	NSE	PBIAS (%)	RMSE	MAE
Random Forest	0.78	−19.17	1.08	0.61
Gradient Boosting	0.78	−18.00	1.06	0.59
Bagging	0.76	−19.71	1.11	0.62
AdaBoost	0.75	−31.13	1.13	0.74

4.3.2. Runoff Reconstruction of the Kalakash River

Plotting the monthly measured runoff curves of the Kalakash River from 1958 to 1998 (Figure 10a), the fluctuation of the runoff curves with the alternation of seasons is regular and shows a cyclic trend. Reconstruction of the time series curves of the Kalakash River runoff data from 1999–2019 based on gradient boosting, random forest, AdaBoost, and bagging methods can be seen in Figure 10b. The comparison showed that the monthly runoff reconstruction curve and the measured data curve fluctuation status were basically consistent, the fluctuation pattern was similar, and could well reflect the intra-annual and inter-annual changes in the runoff features of the Kalakash River. The four models reconstructed runoff curves with similar patterns of change and high fit. They all reached the maximum runoff value at the same instant, and the results were credible. It was further shown that the ML method can be successfully applied to reconstruct the runoff sequence in the HCMA and used to predict the timing of future runoff peaks.

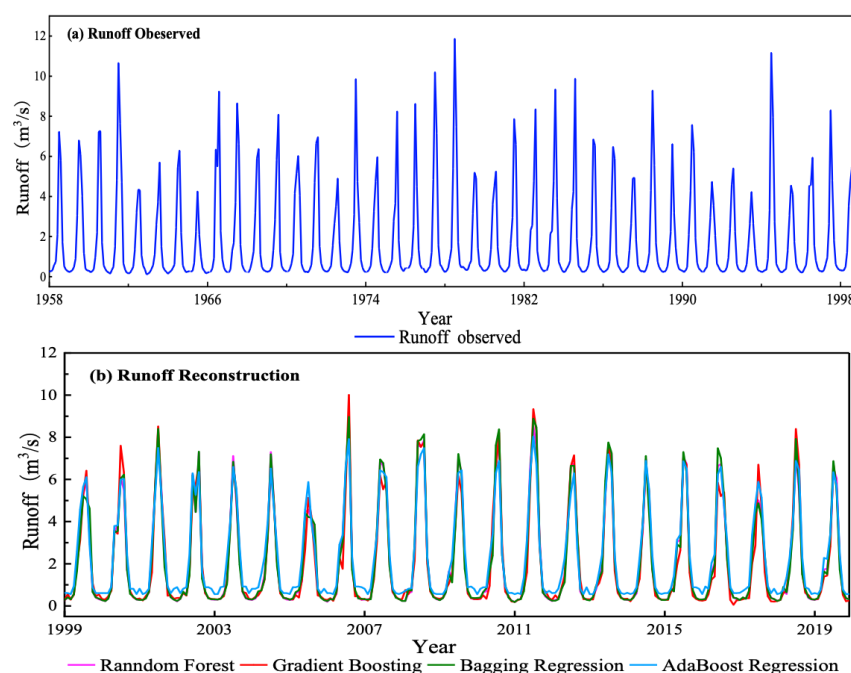


Figure 10. Measured runoff curve of the 1958–1998 data (a). Reconstructed runoff curve of the 1999–2019 data (b).

5. Discussion

Machine learning models have become an important research tool in the field of hydrology [82]. In this paper, we considered the effects of different environmental factors on runoff and applied eight classical machine learning regression models to simulate and reconstruct the monthly runoff of the Yurungkash and Kalakash Rivers in the alpine mountainous region with missing data. Li et al. [83] applied five machine learning models, namely long short-term memory (LSTM), gradient boosted decision tree (GBDT), random forest (RF), SVR, and gate recurrent unit (GRU), and further improved and enhanced them by using stepwise regression with Copula entropy to simulate runoff from two key hydrological stations in the upper Yangtze River in Xinjiang Uygur Autonomous Region, namely Gaochang and Cuntan. For the basin runoff simulation of the two key hydrological stations, NSE was greater than 0.84 with high accuracy, while this paper only used machine learning combined with the regression method for the simulation. Although the results are more ideal, there is still room for progress in improving the method. Wang et al. [84] successfully simulated the daily runoff of snowmelt in the Xiying River Basin in the HCMA of northwest China based on the random forest and artificial neural network (ANN) models, with NSEs of 0.701 and 0.748, respectively. They introduced the remotely sensed snow data and proved that the introduction of the snow data could effectively improve the accuracy of the runoff simulation, and the NSEs of the two models increased by 0.099 and 0.207, respectively. In this paper, we chose to introduce the atmospheric circulation data, and the NSE of the random forest model was 0.82, which is an increase of 0.119 compared with their research accuracy. Han et al. [85] performed a complex parameterization process based on the J2000 physical hydrological model and successfully simulated the runoff of Yamdrok Lake on the Tibetan Plateau of the HCMA, showing high performance with NSE and RMSE values of 0.62 and 1.77, respectively, compared to the present study of gradient boosting (NSE = 0.84, RMSE = 1.24), random forest (NSE = 0.82, RMSE = 1.32), AdaBoost (NSE = 0.78, RMSE = 1.38) and bagging (NSE = 0.78, RMSE = 1.42), where there was a certain gap in the values, which further suggests that the machine learning model is more focused on improving the accuracy of the simulation process compared to the hydrological models.

The input parameters all had varying degrees of influence on the HCMA runoff, with air temperature having the greatest influence and the precipitation correlation being insignificant. Moderate multicollinearity existed between the variables due to the limitation of the sampling capacity and the fixed model; we only removed the relevant variable data by removing the runoff simulation, again, the results found that the NSE did not increase but rather decreased. As there is no method of increasing or decreasing the sampling capacity to eliminate the multicollinearity and judgement, the results of removing the variables indirectly show that ML is able to adapt to the moderate multicollinearity. At the beginning of this study, we compared the results of runoff simulation between the maximum, minimum, and average air and soil temperatures as input parameters and only for the average air and soil temperature data was the NSE value of the former relatively low, and the correlation between the maximum and minimum air and soil temperatures and runoff was not separately assessed, which is one of the shortcomings of this study. A careful comparison of the runoff simulation curves with the measured runoff curves showed that the runoff extremes were inaccurately captured or not captured at all, with uncertainties. After reflection, we analyzed this from a data and modeling point of view. Although the hydrological, meteorological, and atmospheric circulation data at the monthly mean scale have a certain periodicity, due to the complexity of hydrological processes in the HCMA, the runoff sequence becomes nonstationary, and the fluctuation range of the monthly scale data changes drastically, so it is difficult to obtain a strong fit in the runoff simulation, which leads to the influence on the data process of model training in capturing the characteristic extremes and the defects of runoff extreme value capture. Most of the various machine learning models we used simulated runoff based on the default parameters of the model itself without careful parameter tuning, and although the accuracy was high, the capture of runoff peaks was also flawed in certain ways. We used random forest, AdaBoost, and

gradient boost as examples of carefully tuned parameters. By increasing the max_depth of the random forest tree and the value of the random_state parameter, we found that the NSE increased by 0.004. By decreasing the value of the n_estimators parameter in the AdaBoost model, we found that the NSE increased by 0.029, and by increasing the n_estimators and learning_rate parameters in gradient boost, we found that the NSE increased by 0.016, and the results proved that the accuracy of runoff feature capture can be effectively improved by carefully changing the parameters of the model, which is an inspiration for further extension and improvement of the model.

In general, machine learning models outperform univariate models in both the training and testing phases [86]. In this paper, multiple machine learning models were fused and integrated to simulate runoff, and the runoff simulation and reconstruction results corresponding to each method were obtained in only 37 s, which is shorter than the training and prediction time of the LSTM and ANN models ($t = 120\text{--}200$ s) [87]. Before running the simulation, we normalized and standardized the data so that the results were not affected by one or more attributes being too large or too small [88]. Parameterizing was conducted on the pairwise random forest, gradient boost, AdaBoost, and bagging models based on the default parameters (Table A1), specifically: random forest (bootstrap = True, criterion = 'mse', max_depth = 100, max_samples = 490, n_estimators = 1000, random_state = 99), gradient boost (n_estimators = 2000, learning_rate = 0.01, max_depth = 15, max_features = 'sqrt', alpha = 0.9), AdaBoost (n_estimators = 50, learning_rate = 1.0, random_state = None, base_estimator = None, loss = 'linear'), bagging (n_estimators = 90, oob_score = True, random_state = 90, max_samples = 490). Adjusted runoff simulation of the four models showed better performance, relative to the KNN, SVR, ridge, linear regression model performance was excellent, the NSE reached more than 0.78, and could better prevent the emergence of the overfitting phenomenon as well as adapt to the impact of multiple covariance so that the simulation results of the degree of error to reach the minimum. However for the capture of runoff peaks, the accuracy of the simulation needs to be further improved. Linear regression and ridge regression models have a simple model structure, are very sensitive to data, the predicted runoff values appear negative, and the prediction results are unsatisfactory. Hyperparameters have a large impact on the prediction results of the model, and model parameter tuning was weak part of this study. SVR, KNN, and the other four methods choose the default hyperparameters, therefore, the simulation process may show the overfitting and underfitting phenomenon, the model performance is poor, and the simulation accuracy is low, so there is a need to further improve the sample capacity and adjust the hyperparameters to avoid these issues. The ML we used to construct the hydrological model as a black box model is vague in its description of the modeling process, lacks some theoretical support, and is weak in its explanatory nature, which makes it difficult to make meaningful comparisons between different models. The problems of overfitting results and the use of multiple models for feature selection were also not fully addressed.

6. Conclusions

The starting point of this study was to provide better hydrological and water management forecasts for the data scarce HCMA. We combined the meteorological, runoff, and atmospheric circulation data from the northern slopes of the Kunlun Mountains and the Hotan River Basin to construct a time series, completed runoff simulations of the Yurungkash and Kalakash Rivers based on machine learning, and successfully reconstructed the runoff of the Kalakash River for the year with missing data. The main conclusions are:

- (1) Temperature is the most important driver of runoff changes in the mountainous areas upstream of the Hotan River, followed by precipitation, hours of sunshine, wind speed, and weak correlation of atmospheric circulation. The random forest features were ranked in order of importance as $T_mean > RH > DT_mean > Wind > WSPHI > PDO > Sun > NAO > P20_20 > AO > ENSO > AMO$, with a total of 12 elements selected as the machine learning training input data.

- (2) Machine learning (ML) methods can successfully simulate runoff changes in the HCMA. Comprehensive runoff curve coincidence and evaluation parameters using gradient boosting, random forest, AdaBoost, and bagging showed obvious advantages over several other ML methods with NSE of 0.84, 0.82, 0.78, and 0.78, respectively, and the other four methods performed the simulation poorly.
- (3) The four methods including random forest were applied to simulate the runoff of the Kalakash River from 1978 to 1998 with good results, and the Nash–Sutcliffe efficiency coefficients exceeded 0.75. The reconstruction results of the runoff data of the missing period (1999–2019) reflected the intra-annual and inter-annual variations of the runoff characteristics.

Achieving better water resources management on the northern slopes of the Kunlun Mountains has been the direction of research by scientists. In the future, machine learning models can be used to simulate the runoff of the Yarkant and Keriya Rivers on the northern slopes of the Kunlun Mountains to further explore the mechanism of water resource change on the northern slopes of the Kunlun Mountains. The use of less variable daily-scale meteorological and hydrological data as well as remote sensing data of glacier and snowmelt area changes as inputs to the machine learning model can further optimize the runoff simulation and reconstruction results and improve the accuracy of the model. We also found that Khandelwal et al. [89] achieved higher accuracy than machine learning for runoff simulation based on the LSTM framework combined with the SWAT model, suggesting that there is no conflict between traditional “process-oriented” physical hydrological models and “data-oriented” machine learning. The combination of machine learning and physical modeling will lead to more accurate runoff simulation results, which is a new approach that should be considered and used in future hydrological research.

Author Contributions: All authors participated in the study. Formal analysis, S.W. and M.S.; Writing and original draft preparation, S.W.; Data curation, S.W., G.W., Z.X., R.F. and Y.Y.; Conceptualization, S.W. and M.S.; Project administration, S.W., Writing—review and editing, S.W., M.S., X.Y., H.D. and J.L.; Methodology, S.W. and M.S.; Funding acquisition, M.S.; Investigation, M.S. and G.W.; Resources, M.W.; Supervision, M.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (No. 42161027), the Third Xinjiang Scientific Expedition Program (No. 2021xjkk0101), the Third Xinjiang Scientific Expedition Program (No. 2021xjkk0801) and the Natural Science Foundation of Gansu Province (No. 21JR7RA143).

Data Availability Statement: Not applicable.

Acknowledgments: We are grateful to the School of Geography and Environmental Science, Northwest Normal University; the National Natural Science Foundation of China (No. 42161027); the Third Xinjiang Scientific Research Program (No. 2021xjkk0101); the Third Xinjiang Scientific Expedition Program (No. 2021xjkk0801); the Natural Science Foundation of Gansu Province (No. 21JR7RA143) for providing the scientific research platforms and funding. We thank the anonymous reviewers and editorial staff for their constructive and helpful suggestions.

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

Appendix A

Table A1. Hyperparameter adjustment.

ML Models	Parameters
Random Forest	Bootstrap = True, criterion = ‘mse’, max_depth = 100, max_samples = 490, n_estimators = 1000, random_state = 99

Table A1. Cont.

ML Models	Parameters
Gradient Boosting	n_estimators = 2000, learning_rate = 0.01, max_depth = 15, max_features = 'sqrt', alpha = 0.9
SVR	Verbose = False, degree = 3, coef0 = 0.0, kernel = 'rbf', tol = 0.001, epsilon = 0.1, max_iter = -1, shrinking = True, cache_size = 200
AdaBoost	n_estimators = 50, learning_rate = 1.0, random_state = None, base_estimator = None, loss = 'linear'
KNN	n_neighbors = 4, weights = 'uniform', metric_params = None, n_jobs = None, p = 2, algorithm = 'auto'
Bagging	n_estimators = 90, oob_score = True, random_state = 90, max_samples = 490
Ridge	Normalize = False, fit_intercept = True, alpha = 1.0, copy_X = True, max_iter = None, tol = 0.001, solver = 'auto', random_state = None
Linear Regression	fit_intercept = True, normalize = False, copy_X = True, n_jobs = None, positive = False

References

- Luo, M.; Liu, T.; Meng, F.; Duan, Y.; Bao, A.; Xing, W.; Feng, X.; De Maeyer, P.; Frankl, A. Identifying climate change impacts on water resources in Xinjiang, China. *Sci. Total Environ.* **2019**, *676*, 613–626. [\[CrossRef\]](#) [\[PubMed\]](#)
- Rangecroft, S.; Harrison, S.; Anderson, K.; Magrath, J.; Castel, A.P.; Pacheco, P. Climate change and water resources in arid mountains: An example from the Bolivian Andes. *Ambio* **2013**, *42*, 852–863. [\[CrossRef\]](#) [\[PubMed\]](#)
- Zhang, J.; Xu, B.; Gu, Z.; Lv, Y.; Yin, Z.; Guo, X.; Li, L. Coupling of river discharges and alpine glaciers in arid Central Asia. *Quat. Int.* **2023**, *667*, 19–28. [\[CrossRef\]](#)
- Yang, B.; Du, W.; Li, J.; Bao, A.; Ge, W.; Wang, S.; Lyu, X.; Gao, X.; Cheng, X. The Influence of Glacier Mass Balance on River Runoff in the Typical Alpine Basin. *Water* **2023**, *15*, 2762. [\[CrossRef\]](#)
- Wang, C.; Xu, J.; Chen, Y.; Bai, L.; Chen, Z. A hybrid model to assess the impact of climate variability on streamflow for an ungauged mountainous basin. *Clim. Dyn.* **2018**, *50*, 2829–2844. [\[CrossRef\]](#)
- Jiang, J.; Cai, M.; Xu, Y.; Fang, G. The changing trend of flooding in the Aksu River basin. *J. Glaciol. Geocryol* **2021**, *43*, 1–10.
- Wang, X.; Chen, R.; Li, K.; Yang, Y.; Liu, J.; Liu, Z.; Han, C. Trends and Variability in Flood Magnitude: A Case Study of the Floods in the Qilian Mountains, Northwest China. *Atmosphere* **2023**, *14*, 557. [\[CrossRef\]](#)
- Sommer, C.; Malz, P.; Seehaus, T.C.; Lippl, S.; Zemp, M.; Braun, M.H. Rapid glacier retreat and downwasting throughout the European Alps in the early 21st century. *Nat. Commun.* **2020**, *11*, 3209. [\[CrossRef\]](#)
- Wang, J.; Liu, D.W.; Tian, S.N.; Ma, J.L.; Wang, L.X. Coupling reconstruction of atmospheric hydrological profile and dry-up risk prediction in a typical lake basin in arid area of China. *Sci. Rep.* **2022**, *12*, 6535. [\[CrossRef\]](#)
- Mo, K.L.; Chen, Q.W.; Chen, C.; Zhang, J.Y.; Wang, L.; Bao, Z.X. Spatiotemporal variation of correlation between vegetation cover and precipitation in an arid mountain-oasis river basin in northwest China. *J. Hydrol.* **2019**, *574*, 138–147. [\[CrossRef\]](#)
- Yan, L.; Lei, Q.W.; Jiang, C.; Yan, P.T.; Ren, Z.; Liu, B.; Liu, Z.J. Climate-informed monthly runoff prediction model using machine learning and feature importance analysis. *Front. Environ. Sci.* **2022**, *10*, 1049840. [\[CrossRef\]](#)
- Xiao, C.; Zhong, Y.; Wu, Y.; Bai, H.; Li, W.; Wu, D.; Wang, C.; Tian, B. Applying Reconstructed Daily Water Storage and Modified Wetness Index to Flood Monitoring: A Case Study in the Yangtze River Basin. *Remote Sens.* **2023**, *15*, 3192. [\[CrossRef\]](#)
- Wang, F.; Shao, W.; Yu, H.J.; Kan, G.Y.; He, X.Y.; Zhang, D.W.; Ren, M.L.; Wang, G. Re-evaluation of the power of the mann-kendall test for detecting monotonic trends in hydrometeorological time series. *Front. Earth Sci.* **2020**, *8*, 14. [\[CrossRef\]](#)
- Abebe, S.A.; Qin, T.; Zhang, X.; Yan, D. Wavelet transform-based trend analysis of streamflow and precipitation in Upper Blue Nile River basin. *J. Hydrol. Reg. Stud.* **2022**, *44*, 101251. [\[CrossRef\]](#)
- Huang, T.T.; Wang, Z.H.; Wu, Z.Y.; Xiao, P.Q.; Liu, Y. Attribution analysis of runoff evolution in Kuye River Basin based on the time-varying budyko framework. *Front. Earth Sci.* **2023**, *10*, 1092409. [\[CrossRef\]](#)
- Quang, N.H.; Loc, H.H.; Park, E. Characterizing sediment load variability in the red river system using empirical orthogonal function analysis: Implications for water resources management in data poor regions. *J. Hydrol.* **2023**, *624*, 129891. [\[CrossRef\]](#)
- Feng, Z.; Niu, W.; Tang, Z.; Xu, Y.; Zhang, H. Evolutionary artificial intelligence model via cooperation search algorithm and extreme learning machine for multiple scales nonstationary hydrological time series prediction. *J. Hydrol.* **2021**, *595*, 126062. [\[CrossRef\]](#)
- Amiri, S.N.; Khoshravesh, M.; Valashedi, R.N. Assessing the effect of climate and land use changes on the hydrologic regimes in the upstream of Tajan river basin using SWAT model. *Appl. Water Sci.* **2023**, *13*, 130. [\[CrossRef\]](#)

19. Zhao, Y.; Chen, Y.; Zhu, Y.; Xu, S. Evaluating the Feasibility of the Liuxihe Model for Forecasting Inflow Flood to the Fengshuba Reservoir. *Water* **2023**, *15*, 1048. [\[CrossRef\]](#)
20. Liu, J.; Liu, T.; Bao, A.; De Maeyer, P.; Kurban, A.; Chen, X. Response of hydrological processes to input data in high alpine catchment: An assessment of the Yarkant River Basin in China. *Water* **2016**, *8*, 181. [\[CrossRef\]](#)
21. Luo, M.; Meng, F.; Liu, T.; Duan, Y.; Frankl, A.; Kurban, A.; De Maeyer, P. Multi-model ensemble approaches to assessment of effects of local Climate Change on water resources of the Hotan River Basin in Xinjiang, China. *Water* **2017**, *9*, 584. [\[CrossRef\]](#)
22. He, C.; Chen, F.; Long, A.; Qian, Y.; Tang, H. Improving the precision of monthly runoff prediction using the combined non-stationary methods in an oasis irrigation area. *Agric. Water Manag.* **2023**, *279*, 108161. [\[CrossRef\]](#)
23. Perrin, C.; Oudin, L.; Andreassian, V.; Rojas-Serna, C.; Michel, C.; Mathevet, T. Impact of limited streamflow data on the efficiency and the parameters of rainfall—Runoff models. *Hydrol. Sci. J.* **2007**, *52*, 131–151. [\[CrossRef\]](#)
24. Lu, M.; Hou, Q.; Qin, S.; Zhou, L.; Hua, D.; Wang, X.; Cheng, L. A Stacking Ensemble Model of Various Machine Learning Models for Daily Runoff Forecasting. *Water* **2023**, *15*, 1265. [\[CrossRef\]](#)
25. Mohammadi, B. A review on the applications of machine learning for runoff modeling. *Sustain. Water Resour. Manag.* **2021**, *7*, 98. [\[CrossRef\]](#)
26. Hao, R.; Bai, Z. Comparative Study for Daily Streamflow Simulation with Different Machine Learning Methods. *Water* **2023**, *15*, 1179. [\[CrossRef\]](#)
27. Rizeei, H.M.; Pradhan, B.; Saharkhiz, M.A. An integrated fluvial and flash pluvial model using 2D high-resolution sub-grid and particle swarm optimization-based random forest approaches in GIS. *Complex Intell. Syst.* **2019**, *5*, 283–302. [\[CrossRef\]](#)
28. Langhammer, J. Flood Simulations Using a Sensor Network and Support Vector Machine Model. *Water* **2023**, *15*, 2004. [\[CrossRef\]](#)
29. Vaheddoost, B.; Safari, M.J.S.; Yilmaz, M.U. Rainfall-runoff simulation in ungauged tributary streams using drainage area ratio-based multivariate adaptive regression spline and random forest hybrid models. *Pure Appl. Geophys.* **2023**, *180*, 365–382. [\[CrossRef\]](#)
30. Nasiboglu, R.; Nasibov, E. WABL method as a universal defuzzifier in the fuzzy gradient boosting regression model. *Expert Syst. Appl.* **2023**, *212*, 118771. [\[CrossRef\]](#)
31. Noble, W.S. What is a support vector machine? *Nat. Biotechnol.* **2006**, *24*, 1565–1567. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Ratnasingam, S.; Muñoz-Lopez, J. Distance Correlation-Based Feature Selection in Random Forest. *Entropy* **2023**, *25*, 1250. [\[CrossRef\]](#)
33. Natekin, A.; Knoll, A. Gradient boosting machines, a tutorial. *Front. Neurobotics* **2013**, *7*, 21. [\[CrossRef\]](#)
34. Niu, W.-J.; Feng, Z.-K.; Xu, Y.-S.; Feng, B.-F.; Min, Y.-W. Improving prediction accuracy of hydrologic time series by least-squares support vector machine using decomposition reconstruction and swarm intelligence. *J. Hydrol. Eng.* **2021**, *26*, 04021030. [\[CrossRef\]](#)
35. Fu, X.; Shen, B.; Dong, Z.; Zhang, X. Assessing the impacts of changing climate and human activities on streamflow in the Hotan River, China. *J. Water Clim. Chang.* **2020**, *11*, 166–177. [\[CrossRef\]](#)
36. Wei, X.; Long, A.; Yin, Z.; Jiawen, Y. Simulation of response of glacier runoff to climate change in the Hotan River Basin. *Water Resour. Prot.* **2022**, *38*, 137–144.
37. Xu, Y. A study of comprehensive evaluation of the water resource carrying capacity in the arid area: A case study in the Hetian river basin of Xinjiang. *J. Nat. Resour.* **1993**, *8*, 229–237.
38. Fan, M.; Xu, J.; Chen, Y.; Li, W. Modeling streamflow driven by climate change in data-scarce mountainous basins. *Sci. Total Environ.* **2021**, *790*, 148256. [\[CrossRef\]](#)
39. Wang, X.; Luo, Y.; Sun, L.; Shafeeqe, M. Different climate factors contributing for runoff increases in the high glacierized tributaries of Tarim River Basin, China. *J. Hydrol. Reg. Stud.* **2021**, *36*, 100845. [\[CrossRef\]](#)
40. Guo, H.; Ling, H.; Xu, H.; Guo, B. Study of suitable oasis scales based on water resource availability in an arid region of China: A case study of Hotan River Basin. *Environ. Earth Sci.* **2016**, *75*, 984. [\[CrossRef\]](#)
41. Tan, K.; Wang, X.; Gao, H. Analysis of ecological effects of comprehensive treatment in the Tarim River Basin using remote sensing data. *Min. Sci. Technol.* **2011**, *21*, 519–524. [\[CrossRef\]](#)
42. Xue, X.; Mi, Y.; Li, Z.; Chen, Y. Long-term trends and sustainability analysis of air temperature and precipitation in the Hotan River Basin. *Resour. Sci.* **2008**, *30*, 1833–1838.
43. Luo, M.; Liu, T.; Meng, F.; Duan, Y.; Huang, Y.; Frankl, A.; De Maeyer, P. Proportional coefficient method applied to TRMM rainfall data: Case study of hydrological simulations of the Hotan River Basin (China). *J. Water Clim. Chang.* **2017**, *8*, 627–640. [\[CrossRef\]](#)
44. Liu, P.; Jiang, Z.; Li, Y.; Lan, F.; Sun, Y.; Yue, X. Quantitative Study on Improved Budyko-Based Separation of Climate and Ecological Restoration of Runoff and Sediment Yield in Nandong Underground River System. *Water* **2023**, *15*, 1263. [\[CrossRef\]](#)
45. Nuber, S.; Rae, J.W.; Zhang, X.; Andersen, M.B.; Dumont, M.D.; Mithan, H.T.; Sun, Y.; De Boer, B.; Hall, I.R.; Barker, S. Indian Ocean salinity build-up primes deglacial ocean circulation recovery. *Nature* **2023**, *617*, 306–311. [\[CrossRef\]](#) [\[PubMed\]](#)
46. Ye, Y.; Li, Z.; Li, X.; Li, Z. Projection and Analysis of Floods in the Upper Heihe River Basin under Climate Change. *Atmosphere* **2023**, *14*, 1083. [\[CrossRef\]](#)
47. Zhang, Q.; Shen, Z.; Pokhrel, Y.; Farinotti, D.; Singh, V.P.; Xu, C.-Y.; Wu, W.; Wang, G. Oceanic climate changes threaten the sustainability of Asia's water tower. *Nature* **2023**, *615*, 87–93. [\[CrossRef\]](#)
48. Wang, J.; Sun, X.; Cheng, Q.; Cui, Q. An innovative random forest-based nonlinear ensemble paradigm of improved feature extraction and deep learning for carbon price forecasting. *Sci. Total Environ.* **2021**, *762*, 143099. [\[CrossRef\]](#) [\[PubMed\]](#)

49. Shah, S.H.; Angel, Y.; Houborg, R.; Ali, S.; McCabe, M.F. A random forest machine learning approach for the retrieval of leaf chlorophyll content in wheat. *Remote Sens.* **2019**, *11*, 920. [\[CrossRef\]](#)
50. Nguyen, J.M.; Jézéquel, P.; Gillois, P.; Silva, L.; Ben Azzouz, F.; Lambert-Lacroix, S.; Juin, P.; Campone, M.; Gaultier, A.; Moreau-Gaudry, A. Random forest of perfect trees: Concept, performance, applications and perspectives. *Bioinformatics* **2021**, *37*, 2165–2174. [\[CrossRef\]](#)
51. Saravanan, S.; Abijith, D.; Reddy, N.M.; Parthasarathy, K.; Janardhanam, N.; Sathiyamurthi, S.; Sivakumar, V. Flood susceptibility mapping using machine learning boosting algorithms techniques in Idukki district of Kerala India. *Urban Clim.* **2023**, *49*, 101503. [\[CrossRef\]](#)
52. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [\[CrossRef\]](#)
53. Kurani, A.; Doshi, P.; Vakharia, A.; Shah, M. A comprehensive comparative study of artificial neural network (ANN) and support vector machines (SVM) on stock forecasting. *Ann. Data Sci.* **2023**, *10*, 183–208. [\[CrossRef\]](#)
54. Zhong, W.; Du, L. Predicting Traffic Casualties Using Support Vector Machines with Heuristic Algorithms: A Study Based on Collision Data of Urban Roads. *Sustainability* **2023**, *15*, 2944. [\[CrossRef\]](#)
55. Ren, J.; Zhao, H.; Zhang, L.; Zhao, Z.; Xu, Y.; Cheng, Y.; Wang, M.; Chen, J.; Wang, J. Design optimization of cement grouting material based on adaptive boosting algorithm and simplicial homology global optimization. *J. Build. Eng.* **2022**, *49*, 104049. [\[CrossRef\]](#)
56. Wang, C.; Xu, S.; Yang, J. Adaboost algorithm in artificial intelligence for optimizing the IRI prediction accuracy of asphalt concrete pavement. *Sensors* **2021**, *21*, 5682. [\[CrossRef\]](#) [\[PubMed\]](#)
57. Uddin, S.; Haque, I.; Lu, H.; Moni, M.A.; Gide, E. Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction. *Sci. Rep.* **2022**, *12*, 6256. [\[CrossRef\]](#)
58. Wang, F.; Zhen, Z.; Wang, B.; Mi, Z. Comparative study on KNN and SVM based weather classification models for day ahead short term solar PV power forecasting. *Appl. Sci.* **2017**, *8*, 28. [\[CrossRef\]](#)
59. Garcia, S.; Derrac, J.; Cano, J.; Herrera, F. Prototype selection for nearest neighbor classification: Taxonomy and empirical study. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 417–435. [\[CrossRef\]](#) [\[PubMed\]](#)
60. Rajan, M. An efficient Ridge regression algorithm with parameter estimation for data analysis in machine learning. *SN Comput. Sci.* **2022**, *3*, 171. [\[CrossRef\]](#)
61. Wang, X.; Wang, X.; Ma, B.; Li, Q.; Wang, C.; Shi, Y. High-performance reversible data hiding based on ridge regression prediction algorithm. *Signal Process.* **2023**, *204*, 108818. [\[CrossRef\]](#)
62. Hothorn, T.; Lausen, B. Double-bagging: Combining classifiers by bootstrap aggregation. *Pattern Recognit.* **2003**, *36*, 1303–1309. [\[CrossRef\]](#)
63. Wang, Q.; Luo, Z.; Huang, J.; Feng, Y.; Liu, Z. A novel ensemble method for imbalanced data learning: Bagging of extrapolation-SMOTE SVM. *Comput. Intell. Neurosci.* **2017**, *2017*, 1827016. [\[CrossRef\]](#) [\[PubMed\]](#)
64. James, G.; Witten, D.; Hastie, T.; Tibshirani, R.; Taylor, J. Linear regression. In *An Introduction to Statistical Learning: With Applications in Python*; Springer: Berlin/Heidelberg, Germany, 2023; pp. 69–134.
65. Parashar, A.; Parashar, A.; Ding, W.; Shabaz, M.; Rida, I. Data Preprocessing and Feature Selection Techniques in Gait Recognition: A Comparative Study of Machine Learning and Deep Learning Approaches. *Pattern Recognit. Lett.* **2023**, *172*, 65–73. [\[CrossRef\]](#)
66. Wang, W.; Jing, H.; Guo, X.; Dou, B.; Zhang, W. Analysis of Water and Salt Spatio-Temporal Distribution along Irrigation Canals in Ningxia Yellow River Irrigation Area, China. *Sustainability* **2023**, *15*, 12114. [\[CrossRef\]](#)
67. Zhao, Y.; Zhu, W.; Wei, P.; Fang, P.; Zhang, X.; Yan, N.; Liu, W.; Zhao, H.; Wu, Q. Classification of Zambian grasslands using random forest feature importance selection during the optimal phenological period. *Ecol. Indic.* **2022**, *135*, 108529. [\[CrossRef\]](#)
68. Alduailij, M.; Khan, Q.W.; Tahir, M.; Sardaraz, M.; Alduailij, M.; Malik, F. Machine-learning-based DDoS attack detection using mutual information and random forest feature importance method. *Symmetry* **2022**, *14*, 1095. [\[CrossRef\]](#)
69. Fu, H.; Shen, Y.; Liu, J.; He, G.; Chen, J.; Liu, P.; Qian, J.; Li, J. Cloud detection for FY meteorology satellite based on ensemble thresholds and random forests approach. *Remote Sens.* **2018**, *11*, 44. [\[CrossRef\]](#)
70. Han, H.; Morrison, R.R. Improved runoff forecasting performance through error predictions using a deep-learning approach. *J. Hydrol.* **2022**, *608*, 127653. [\[CrossRef\]](#)
71. Vu, M.; Raghavan, S.V.; Liong, S.-Y. SWAT use of gridded observations for simulating runoff—a Vietnam river basin study. *Hydrol. Earth Syst. Sci.* **2012**, *16*, 2801–2811. [\[CrossRef\]](#)
72. Lu, X.; Li, J.; Liu, Y.; Li, Y.; Huo, H. Quantitative Precipitation Estimation in the Tianshan Mountains Based on Machine Learning. *Remote Sens.* **2023**, *15*, 3962. [\[CrossRef\]](#)
73. Yapo, P.O.; Gupta, H.V.; Sorooshian, S. Automatic calibration of conceptual rainfall-runoff models: Sensitivity to calibration data. *J. Hydrol.* **1996**, *181*, 23–48. [\[CrossRef\]](#)
74. Gu, X.; Yang, G.; He, X.; Zhao, L.; Li, X.; Li, P.; Liu, B.; Gao, Y.; Xue, L.; Long, A. Hydrological process simulation in Manas River Basin using CMADS. *Open Geosci.* **2020**, *12*, 946–957. [\[CrossRef\]](#)
75. Lee, J.; Noh, J. Development of a One-Parameter New Exponential (ONE) Model for Simulating Rainfall-Runoff and Comparison with Data-Driven LSTM Model. *Water* **2023**, *15*, 1036. [\[CrossRef\]](#)
76. Xu, Y.; Hu, C.; Wu, Q.; Jian, S.; Li, Z.; Chen, Y.; Zhang, G.; Zhang, Z.; Wang, S. Research on particle swarm optimization in LSTM neural networks for rainfall-runoff simulation. *J. Hydrol.* **2022**, *608*, 127553. [\[CrossRef\]](#)

77. Feng, Z.-K.; Niu, W.-J.; Wan, X.-Y.; Xu, B.; Zhu, F.-L.; Chen, J. Hydrological time series forecasting via signal decomposition and twin support vector machine using cooperation search algorithm for parameter identification. *J. Hydrol.* **2022**, *612*, 128213. [[CrossRef](#)]
78. Patro, E.R.; De Michele, C.; Avanzi, F. Future perspectives of run-of-the-river hydropower and the impact of glaciers' shrinkage: The case of Italian Alps. *Appl. Energy* **2018**, *231*, 699–713. [[CrossRef](#)]
79. Taylor, G.P.; Loikith, P.C.; Aragon, C.M.; Lee, H.; Waliser, D.E. CMIP6 model fidelity at simulating large-scale atmospheric circulation patterns and associated temperature and precipitation over the Pacific Northwest. *Clim. Dyn.* **2023**, *60*, 2199–2218. [[CrossRef](#)]
80. Fahu, C.; Tingting, X.; Yujie, Y.; Shengqian, C.; Feng, C.; Wei, H.; Jie, C. Discussion on the problem of “warming and humidification” and its future trend in the arid area of Northwest China. *Sci. China Earth Sci.* **2023**, *53*, 1246–1262. [[CrossRef](#)]
81. Zhao, Q.; Ye, B.; Ding, Y.; Zhang, S.; Yi, S.; Wang, J.; Shangguan, D.; Zhao, C.; Han, H. Coupling a glacier melt model to the Variable Infiltration Capacity (VIC) model for hydrological modeling in north-western China. *Environ. Earth Sci.* **2013**, *68*, 87–101. [[CrossRef](#)]
82. El Bilali, A.; Abdeslam, T.; Ayoub, N.; Lamane, H.; Ezzaouini, M.A.; Elbeltagi, A. An interpretable machine learning approach based on DNN, SVR, Extra Tree, and XGBoost models for predicting daily pan evaporation. *J. Environ. Manag.* **2023**, *327*, 116890. [[CrossRef](#)] [[PubMed](#)]
83. Li, X.; Zhang, L.; Zeng, S.; Tang, Z.; Liu, L.; Zhang, Q.; Tang, Z.; Hua, X. Predicting Monthly Runoff of the Upper Yangtze River Based on Multiple Machine Learning Models. *Sustainability* **2022**, *14*, 11149. [[CrossRef](#)]
84. Wang, G.; Hao, X.; Yao, X.; Wang, J.; Li, H.; Chen, R.; Liu, Z. Simulations of Snowmelt Runoff in a High-Altitude Mountainous Area Based on Big Data and Machine Learning Models: Taking the Xiyang River Basin as an Example. *Remote Sens.* **2023**, *15*, 1118. [[CrossRef](#)]
85. Tang, H.; Zhang, F.; Zeng, C.; Wang, L.; Zhang, H.; Xiang, Y.; Yu, Z. Simulation of Runoff through Improved Precipitation: The Case of Yamzho Yumco Lake in the Tibetan Plateau. *Water* **2023**, *15*, 490. [[CrossRef](#)]
86. Aksan, F.; Suresh, V.; Janik, P.; Sikorski, T. Load Forecasting for the Laser Metal Processing Industry Using VMD and Hybrid Deep Learning Models. *Energies* **2023**, *16*, 5381. [[CrossRef](#)]
87. Guo, J.; Liu, Y.; Zou, Q.; Ye, L.; Zhu, S.; Zhang, H. Study on optimization and combination strategy of multiple daily runoff prediction models coupled with physical mechanism and LSTM. *J. Hydrol.* **2023**, *624*, 129969. [[CrossRef](#)]
88. Jin, Q.; Sun, Y.; Liu, Z.; He, S. Multidimensional tensor strategy for the inverse analysis of in-service bridge based on SHM data. *Innov. Infrastruct. Solut.* **2023**, *8*, 228. [[CrossRef](#)]
89. Khandelwal, A.; Xu, S.; Li, X.; Jia, X.; Stienbach, M.; Duffy, C.; Nieber, J.; Kumar, V. Physics guided machine learning methods for hydrology. *arXiv* **2020**, arXiv:2012.02854. [[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.