frontiers | Research Topics

Hydroclimatic extremes: Human-natural system adaptation and impacts

Edited by
Chuanfu Zang, Zhenzhong Zeng, Fei Tian, Patrick Laux and Ganquan Mao

Published in
Frontiers in Environmental Science
Frontiers in Earth Science

ISSN 1664-8714
ISBN 978-2-83251-881-6
DOI 10.3389/978-2-83251-881-6

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Hydroclimatic extremes: Human-natural system adaptation and impacts

Topic editors

Ganquan Mao — Southern University of Science and Technology, China
Chuanfu Zang — South China Normal University, China
Zhenzhong Zeng — Southern University of Science and Technology, China
Fei Tian — China Agricultural University, China
Patrick Laux — Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

Citation

Mao, G., Zang, C., Zeng, Z., Tian, F., Laux, P., eds. (2023). *Hydroclimatic extremes: Human-natural system adaptation and impacts*.
Lausanne: Frontiers Media SA. doi: 10.3389/978-2-83251-881-6

Table of contents

frontiers in Environmental Science

ORIGINAL RESEARCH
published: 10 December 2021
doi: 10.3389/fenvs.2021.791336

Estimation of Surface and Near-Surface Air Temperatures in Arid Northwest China Using Landsat Satellite Images

Yi Liu[1], *Samuel Ortega-Farías*[2], *Fei Tian*[1], *Sufen Wang*[1]* *and Sien Li*[1]

[1]Center for Agricultural Water Research in China, China Agricultural University, Beijing, China, [2]Research and Extension Center for Irrigation and Agroclimatology (CITRA) and Research Program on Adaptation of Agriculture to Climate Change (A2C2), Faculty of Agricultural Sciences, Universidad de Talca, Talca, Chile

Near-surface air (Ta) and land surface (Ts) temperatures are essential parameters for research in the fields of agriculture, hydrology, and ecological changes, which require accurate datasets with different temporal and spatial resolutions. However, the sparse spatial distribution of meteorological stations in Northwest China may not effectively provide high-precision Ta data. And it is not clear whether it is necessary to improve the accuracy of Ts which has the most influence on Ta. In response to this situation, the main objective of this study is to estimate Ta for Northwest China using multiple linear regression models (MLR) and random forest (RF) algorithms, based on Landsat 8 images and auxiliary data collected from 2014 to 2019. Ts, NDVI (Normalized Difference Vegetation Index), surface albedo, elevation, wind speed, and Julian day were variables to be selected, then used to estimate the daily average Ta after analysis and adjustment. Also, the Radiative Transfer Equation (RTE) method for calculating Ts would be corrected by NDVI (RTE-NDVI). The results show that: 1) The accuracy of the surface temperature (Ts) was improved by using RTE-NDVI; 2) Both MLR and RF models are suitable for estimating Ta in areas with few meteorological stations; 3) Analyzing the temporal and spatial distribution of errors, it is found that the MLR model performs well in spring and summer, and is lower in autumn, and the accuracy is higher in plain areas away from mountains than in mountainous areas and nearby areas. This study shows that through appropriate selection and combination of variables, the accuracy of estimating the pixel-scale Ta from satellite remote sensing data can be improved in the area that has less meteorological data.

Keywords: average air temperature, land surface temperature, remote sensing, Landsat 8, statistical models

OPEN ACCESS

Edited by:
Tianjie Zhao,
Aerospace Information Research Institute (CAS), China

Reviewed by:
Osman Orhan,
Mersin University, Turkey
Hongbo Su,
Florida Atlantic University, United States

***Correspondence:**
Sufen Wang
wwwsf71@163.com

Specialty section:
This article was submitted to Environmental Informatics and Remote Sensing, a section of the journal Frontiers in Environmental Science

Received: *08 October 2021*
Accepted: *22 November 2021*
Published: *10 December 2021*

Citation:
Liu Y, Ortega-Farías S, Tian F, Wang S and Li S (2021) Estimation of Surface and Near-Surface Air Temperatures in Arid Northwest China Using Landsat Satellite Images. Front. Environ. Sci. 9:791336. doi: 10.3389/fenvs.2021.791336

INTRODUCTION

Near-surface air temperature (Ta), usually refers to the air temperature at 2 m above the ground, is an essential factor affecting ecology, agriculture, and urban areas (Raja Reddy et al., 1997; Krüger and Emmanuel, 2013; Shamir and Georgakakos, 2014), and is also the basis for climate change studies (Alkama and Cescatti, 2016; Bathiany et al., 2018). The traditional method of obtaining Ta mainly relies on the temperature sensor installed at the meteorological station, and the interpolation method is frequently used to extend to regional-scale applications (Mostovoy et al., 2006). If the

meteorological stations are sparse and unevenly distributed, the accuracy of the interpolation method will be greatly restricted due to the influence of underlying surface heterogeneity and heat conduction unevenness to air temperature (Chen et al., 2015). The air temperature has become the primary driving variable of many land surface models (Nieto et al., 2011), so its spatial fidelity must be higher than that obtained by interpolation of point observation data, and even the most complex geostatistical techniques cannot meet the requirements (Prince et al., 1998).

Satellite data has the characteristics of continuous spatial coverage (Czajkowski et al., 1997), which can obtain large-scale atmospheric information and invert surface parameters, including global surface temperature, vegetation index, elevation, and other information (Wang et al., 2018; Yoo et al., 2018). Due to the complexity of atmospheric radiation and its low proportion in remote sensing signals, Ta cannot be directly reversed (Xu et al., 2012; Li and Zha, 2018). However, Ta can be obtained by establishing the regression relationship between Ta and remote sensing inversion and auxiliary parameters, such as using land surface temperature (Ts), Normalized Difference Vegetation Index (NDVI), wind speed, geographic location and elevation, among which Ts is the most important parameter (Vancutsem et al., 2010; Hachem et al., 2012; Song and Wu, 2018).

Ts as the direct driving force of long-wave radiation and turbulent heat flux exchange at the surface-atmosphere interface is one of the most significant parameters in the physical process of surface energy and water balance at regional and global scales (Anderson et al., 2008; Li et al., 2013; Orhan et al., 2014; Folland et al., 2018). At present, there are three main methods for using Landsat to retrieve Ts: atmospheric correction method (also known as Radiative Transfer Equation: RTE), Single Channel Algorithm (SCA), and Split Window Algorithm (SWA). The SWA does not require any atmospheric profile information at the time of collection, which needs to use two thermal infrared channels (Li et al., 2013). However, the United States Geological Survey has pointed out that Thermal Infrared Sensor (TIRS) 11th band has data reception abnormalities and calibration instability problems, which mainly affects the accuracy of the split window algorithm applied to Landsat-8 TIRS data to retrieve Ts (Xu, 2015). SCA and RTE rely on atmospheric transmissivity and upwelling and downwelling atmospheric radiances (Jimenez-Munoz et al., 2009; Sekertekin and Bonafoni, 2020). RTE removes the error caused by the atmosphere's thermal radiation on the surface and converts the thermal radiation intensity to the corresponding Ts (Ma and Pu, 2020). When using different data sets, the performance of the RTE method to retrieve Ts is also different. The Ts calculation result for Landsat TM 5 data is better than other methods in the same period with RMSE is 1–3°C (Sobrino et al., 2004; Ndossi and Avdan, 2016; Windahl and Beurs, 2016); however, the RMSE calculated based on Landsat 8 TIRS 10th band was 1.5–5°C, and most of them are inferior to other algorithms (Ndossi and Avdan, 2016; Wang et al., 2016; Sekertekin and Bonafoni, 2020). Moreover, the overestimation shown in the TIRS band will increase as the proportion of vegetation decreases (Xu and Huang, 2016).

Several studies have used surface information to estimate air temperatures, such as the temperature-vegetation index (TVX), energy balance, statistics, and machine learning methods (Zakšek and Schroedter-Homscheidt, 2009; Benali et al., 2012). Nemani and Running (1989), Goward et al. (1994) proposed the TVX approach to estimate near surface air temperature with promising results. The method is based on the assumption that there is a strong negative correlation between Ts and vegetation index (Goward et al., 1994; Czajkowski et al., 1997). Assuming that the Ta for fully covered vegetation is close to Ts, the value of full coverage NDVI (NDVImax) can be used to obtain an approximate value of Ta (Stisen et al., 2007; Nieto et al., 2011; Zhu et al., 2013). However, this assumption does not apply to all seasons, soil moisture, and ecosystem types, so estimation of Ta by using TVX method is not feasible in areas or seasons without high vegetation cover (Vancutsem et al., 2010).

The energy balance method has a physical mechanism, so it has well portability and versatility (Hou et al., 2013; Shen et al., 2020). According to the energy balance equation, Ta is related to surface temperature, and it depends on various environmental factors such as solar radiation, cloud cover, wind speed, soil moisture, and surface type (Prince et al., 1998). A large number of required parameters cannot be completely retrieved by remote sensing (Mostovoy et al., 2006), so it is difficult to use remote sensing to perform Ta inversion in the area. The issue of unclosed surface energy balance also brings additional uncertainty to this method (Zhang et al., 2015).

Statistical methods need to analyze the relationship between Ta and Ts and other auxiliary data, and then build an estimation model based on specific correlation (Cresswell et al., 1999; Park, 2011), which including simple statistical models, multiple linear regression (MLR) models, geographically weighted regression (GWR) models, and machine learning methods (Vogt et al., 1997; Vancutsem et al., 2010; Shen et al., 2020). Studies have shown that the linear regression models are more accurate in calculating the average daily temperature with a root mean square error (RMSE) ranging between 1.29–3.60°C (Chen et al., 2015; Shi et al., 2016; Yang et al., 2017), which can produce good results in a specific space and time range, but require a large amount of data involved in the calculation and training of the algorithms (Stisen et al., 2007). Geographically weighted statistical and machine learning methods usually have higher accuracy (Moser et al., 2015; Wang et al., 2017, 2018). Geographically and temporally weighted regression (GTWR) is an extension of the general linear regression model, which embeds changes in location and time into the regression equation and estimates regression coefficients for spatio-temporal variation by performing local regressions that can solve for constant-coefficient limits (Bai et al., 2016; Li et al., 2018). Machine learning methods can handle non-linear and highly correlated predictors (James et al., 2013) and estimate the temperature in areas with complex and heterogeneous underlying surfaces, mainly including neural networks (Jang et al., 2004), M5 model trees (Emamifar et al., 2013), and random forests (Zhang et al., 2016), support vector machine (Moser et al., 2015). Random forests (RF) are widely used and have been verified in various terrains. Ho et al. (2014) indicated that the RF algorithm is very useful for mapping the variability of urban internal temperature. Meyer et al. (2016) have pointed out that compared with linear regression, generalized augmented regression model (GBM), and cubic regression, the RF algorithm performs poorly in the

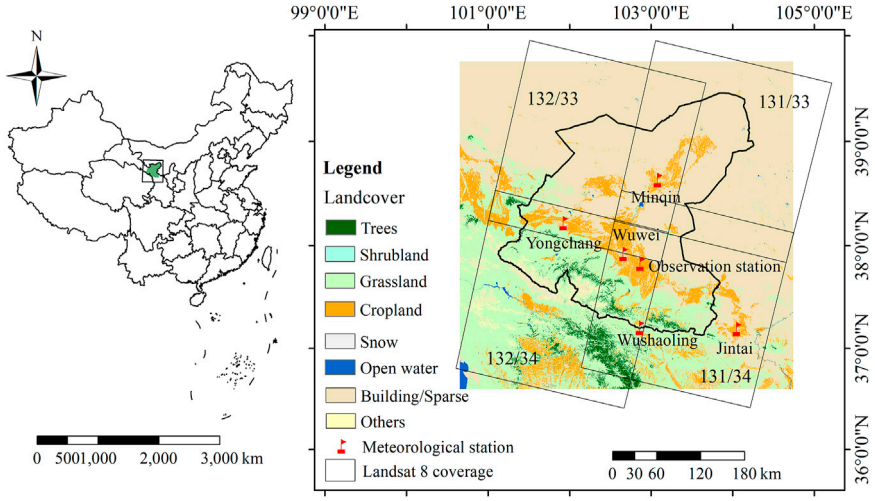


FIGURE 1 | The location of the study area, the coverage area of Landsat8, distribution of meteorological stations, and land use cover.

TABLE 1 | Type and source of data used.

Data	Unit	Time scale	Type	Source
Ta Wind speed	°C m/s	Daily 2014-2019	Points	National Meteorological Information Center (http://data.cma.cn/)
Ts by RTE Albedo	°C	Instantaneous 2014-2019	Grids with a resolution of 30 m × 30 m	"USGS Landsat 8 Surface Reflectance Tier 1" from the GEE platform
NDVI	-			ID: "LANDSAT/LC08/C01/T1_SR"
Elevation	m	2014-2019	Grids with a resolution of 30 m × 30 m	NASA SRTM Digital Elevation 30m from the GEE platform ID: "USGS/SRTMGL1_003"
Surface temperature	°C	Quarter of an hour 2014-2015	Points	SI-111 thermal infrared radiometer Coordinate: 102.885090, 37.824715 (2014) 102.884864, 37.819561 (2015) 102.876055,37.823513 (2016-2018)
Landcover	-	One year 2020	Grids with a resolution of 10 m × 10 m	"ESA WorldCover 10 m v100" from the GEE platform ID: "ESA/WorldCover/v100"

extremely cold Antarctica. However, Noi et al. (2017) have shown reliable results in mountainous areas. Therefore, the conclusion of RF estimation Ta still needs to be discussed.

Most of the current researches for estimating Ta is based on the MODIS due to its time continuity advantage, but its resolution cannot meet the demand for farmland or even smaller scales. The spatial resolution of Landsat 8 imagery is higher than that of MODIS, and the estimated Ta data have a more intuitive correspondence with land use. Generally, if the temperature estimate based on remote sensing data is accurate, the accuracy should be between 1 and 2°C (Vázquez et al., 1997). Research suggests that site density is positively correlated with model accuracy. In other words, the denser the sites, the higher the accuracy of the model (Shen et al., 2020). If Ta is estimated jointly in Northwest China, where meteorological stations are scarce, and in areas where other stations are d ense, the former regions often do not obtain more accurate Ta data (Chen et al., 2015; Li and Zha, 2018). Therefore, it is very necessary to separately model and estimate the area where the data is lacking.

The main objective of this study was to propose a statistical method based on Landsat 8 and auxiliary data for accurately estimating Ta, especially in arid northwest China, where meteorological stations are scarce and unevenly distributed. The specific objectives of this study were to 1) use the RTE method to estimate Ts based on the Landsat8 data, and improve accuracy. 2) Select reasonable independent variables, participate in the modeling of MLR and RF, and compare the results of Ta estimation. 3) Evaluate the performance of the optimal Ta estimation model in time and space scale.

MATERIALS AND METHODOLOGY

Study Region and Meteorological Station

The study area is the Shiyang River basin in the arid region of Northwest China, which is located in the Hexi Corridor of Gansu Province and the coordinate range is between 101°40′E-104°20′E and 36°30′N-39°30′ N (**Figure 1**). It covers an area of 41,600 km^2, and the range of altitude is 1,157m–5012 m. This area is of a continental temperate arid climate, with the characteristics of aridity where precipitation is 300 mm per year and average annual pan evaporation is 2,000 mm.

Meteorological stations in Northwest China are sparse, with only five stations in and near the study area (Minqin, Wuwei, Yongchang, Wushaoling, Jintai station), which can be collected from the China Meteorological Science Data Center (http://data.cma.cn). The observation station (National Field Scientific Observation and Research Station on Efficient Water Use of Oasis Agriculture in Wuwei of Gansu Province) has a meteorological station to collect Ta data, SI-111 thermal infrared radiometers to obtain Ts data (the coordinates in **Table 1**). Google Earth Pro was used to verify the location of the meteorological station and ensure that Ta data corresponds totally to the site location. Then obtain the available satellite images from 2014 to 2019 and calculate and extract the Ts data corresponding to the location of the meteorological station through the GEE platform.

Satellite Data and Processing

Google Earth Engine (GEE) is a geospatial processing platform based on cloud computing developed by Google, which promotes fast analysis by using Google's computing infrastructure and providing a convenient platform for applications based on linking to the cloud computing engine (Becker et al., 2021). Most of the algorithms built into the GEE cloud computing platform use pixel-by-pixel calculation functions, so no matter what the area or proportion of the calculation and analysis is required, as long as the research area has available data. The platform is suitable for scientific researchers with a background in non-professional programming and can quickly realize global-scale remote sensing data processing and mining. This research uses the online JavaScript API of the GEE platform (https://earthengine.google.com/) to access and analyze the data sets used from the public catalog, without downloading images, only outputting the processing results, which improves computing efficiency.

This study uses the Landsat 8 Raw and SR (Surface Reflectance product) dataset in the GEE platform, screening the images with cloudiness not exceeding 30% from 2014 to 2019 and perform cloud removal processing. To cover the entire area of the Shiyang River basin using images of four Landsat-8 tiles 131/033,131/034 and 132/033,132/034 (**Figure 1**). By using the GEE platform, the remote sensing images were spliced efficiently, clipped, and parameter inversion, also the meteorological data were interpolated.

Land Surface Temperature

At present, there are three main methods for remote sensing to retrieve land surface temperature: atmospheric Radiative Transfer Equation (RTE) method, single-channel algorithm, and split-window algorithm. This study uses the Landsat 8 SR data set in the GEE platform, to retrieve the Ts based on the RTE method.

The expression of the radiation transfer equation of the thermal infrared radiation value (L_λ) received by the satellite sensor is (Li et al., 2013; Windahl and Beurs, 2016):

$$L_\lambda = [\varepsilon B(T_s) + (1-\varepsilon)L\downarrow]\tau + L\uparrow \tag{1}$$

where L_λ is the spectral radiance value at the top of the atmosphere at the band λ ($W \cdot m^{-2} \cdot \mu m^{-1} \cdot sr^{-1}$); ε is the surface specific emissivity; B (T_s) is the blackbody thermal emissivity brightness ($W \cdot m^{-2} \cdot \mu m^{-1} \cdot sr^{-1}$); τ is the atmospheric thermal infrared band transmittance; L↓ is the downward radiance of the atmosphere after reflection on the ground ($W \cdot m^{-2} \cdot \mu m^{-1} \cdot sr^{-1}$); L↑ is the upward radiance of the atmosphere ($W \cdot m^{-2} \cdot \mu m^{-1} \cdot sr^{-1}$).

Knowledge of land surface emissivity (LSE) is necessary to apply the above methods to a Landsat image. Considering different situations, obtain the emissivity value from NDVI (Sobrino et al., 2004; Orhan and Yakar, 2016):

$$\varepsilon = \begin{cases} 0.986 & \mathrm{NDVI} \le \mathrm{NDVI}_{Soil} \\ 0.004\left[\dfrac{\mathrm{NDVI} - \mathrm{NDVI}_{Soil}}{\mathrm{NDVI}_{Veg} - \mathrm{NDVI}_{Soil}}\right] + 0.986 & \mathrm{NDVI}_{Soil} < NDVI < \mathrm{NDVI}_{Veg} \\ 0.990 & \mathrm{NDVI} \ge \mathrm{NDVI}_{Veg} \end{cases} \tag{2}$$

where $\mathrm{NDVI}_{Veg} = 0.7$ and $\mathrm{NDVI}_{Soil} = 0.05$ (Ma and Pu, 2020).

When using Landsat8 images, take the 10th band to provide the thermal infrared radiance value. The calculation formula of $B(T_s)$ is:

$$B(T_s) = \frac{[L_\lambda - L\uparrow - (1-\varepsilon)\tau L\downarrow]}{\tau\varepsilon} \tag{3}$$

The calculation of the surface temperature (T_s) uses the Planck formula:

$$T_s = \frac{K_2}{\ln\left(\frac{K_1}{B(T_s)} + 1\right)} - 273.15 \tag{4}$$

where Ts is surface temperature (°C); K_1, K_2 can be obtained from the header file of the remote sensing data. For Landsat8 TIRS Band10, $K_1 = 774.89W \cdot m^{-2} \cdot \mu m^{-1} \cdot sr^{-1}$, $K_2 = 1321.08K$.

It can be seen that the use of Radiative Transfer Equation method to retrieve the Ts needs to have the atmospheric profile parameters, which can be obtained by entering the shadowing time, latitude, and longitude in the website provided by NASA (http://atmcorr.gsfc.nasa.gov/).

Auxiliary Data and Processing Flow

The near-surface air temperature (Ta) has a good correlation with the surface temperature (Ts) (Benali et al., 2012; Ruiz-Álvarez et al., 2019). Ts is a physical quantity that reflects the degree of cold and heat on the surface of a ground object because it is affected by the characteristics of the underlying surface, such as vegetation coverage and dry and wet conditions. Ta is a physical quantity reflecting the degree of cold and hot air in the atmosphere. The atmosphere has strong fluidity, which is easily affected by the surrounding environment (Xu et al., 2012; Gholamnia et al., 2017). Therefore, when looking for the correlation between Ts and Ta, the influence of various factors such as ground characteristic and environment must be considered. For the estimation of Ta, Jang et al. (2004) showed that Julian day is a more significant parameter than altitude or the solar zenith angle. In addition, we have chosen

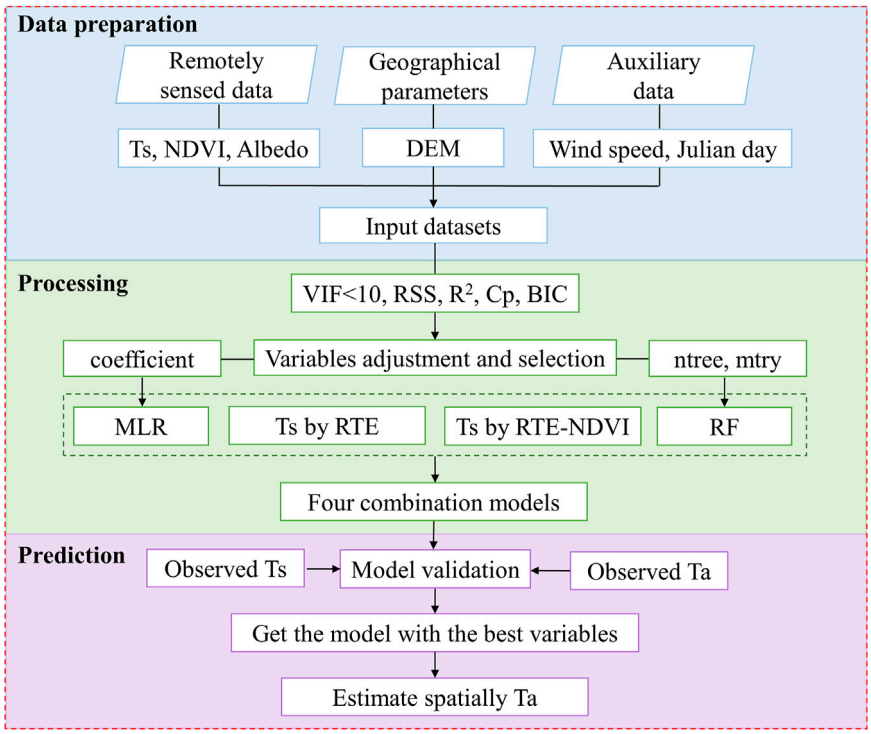


FIGURE 2 | The flow chart for estimating daily mean air temperature (Ta). Input variables include surface temperature (Ts), normalized vegetation index (NDVI), surface albedo (Albedo), elevation (DEM), wind speed, and Julian day, and build MLR and RF models. Variables are selected by indicators such as variance inflation factor (VIF), residual sum of squares (RSS), coefficient of determination (R^2), Mallows Cp (Cp) value, and Bayesian Information Criterion (BIC). Finally, compare the simulation results of the four combined models with the measured values, then obtaining the Ta distribution in the study area.

variables that are often selected as predictors in air temperature modeling literature, such as surface albedo, elevation (DEM), normalized vegetation index (NDVI), and wind speed (Riddering and Queen, 2006; Cristóbal et al., 2008; Hou et al., 2013). The full name of the DEM (WGS84/EGM96) data source is "NASA SRTM Digital Elevation 30 m," which is used directly on the GEE platform and provided by NASA JPL (Farr et al., 2007).

The broadband albedo (α) of the ground surface is a critical variable for many scientific applications, which is the ratio of the total radiant flux reflected by the ground surface to the incident flux (Liang et al., 2003). The calculation formula of α is:

$$\alpha = 0.356\alpha_2 + 0.130\alpha_4 + 0.373\alpha_5 + 0.085\alpha_6 + 0.072\alpha_7 - 0.0018 \quad (5)$$

Where α is the surface reflectance, and its value is between 0–1.0; α_2, α_4, α_5, α_6, and α_7 are the 2, 4, 5, 6, and 7 bands of Landsat 8 surface reflectance products.

As has been commented in the previous section, the LSE can be retrieved from NDVI values. The data can be used to construct NDVI according to the following equation:

$$NDVI = \frac{NIR - R}{NIR + R} \quad (6)$$

Where NIR and R are the reflection values of the near-infrared and infrared bands, respectively, which are the fifth and fourth bands of Landsat 8.

This study mainly considers the relationship between Ta and Ts, NDVI, surface albedo, DEM, wind speed and Julian day. The flow chart of the models used is shown in **Figure 2**, and the process can be summarized as Data preparation, processing and prediction. Data collection is the basis for establishing a reliable model, but it is also necessary to filter the input variables and try to use as few variables as possible under the condition of high model accuracy. Variables used in this study are readily available, which are closely related to the changes of Ta. Previous studies on Ta estimation lack the verification of Ts and analysis of related results. Therefore, this research explored whether it is necessary to improve the precision of Ts to achieve better results in actual application. After verification and comparison, the best model is selected from the combined methods to estimate Ta.

The spatial and temporal matching among the variables is the main issue for the reliability of the regression implementation. **Table 1** shows the sources and resolutions of all data sets. On the spatial scale, the resolutions of spatial variability independent variables such as Ts, Albedo, NDVI, elevation, wind speed (resampled after interpolation) are all

30 m × 30 m, which is spatially consistent. On the time scale, Ts, Albedo, and NDVI are instantaneous data, and only wind speed is daily data.

Models Adjustment and Validation

Multiple Linear Regression

The regression model aims to establish the corresponding functional relationship between several independent input parameters and output targets (Giacomino et al., 2011; Agha and Alnahhal, 2012; Williams and Ojuri, 2021). MLR is a linear regression technique and very useful for the best relationship between predictor variables and several independent variables, which is different from a simple linear regression analysis (Akan et al., 2015). R is an excellent tool for statistical calculation and statistical mapping. It is free, open-source software that does not require any license and is simple to operate (Williams and Ojuri, 2021). Using the "lm" function in R software, an MLR model of Ta and multiple correlation factors are established Expanding the MLR equation into the more commonly used form is:

$$A = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + \ldots + b_nX_n \quad (7)$$

where A is the regression target variable; b_0~b_n are undetermined coefficients; X_1~X_n are independent variables.

Random Forest

The random forest (RF) is a non-parametric machine learning algorithm, which is more flexible than classical statistical models (Genuer et al., 2017; Li et al., 2019). It hardly requires statistical assumptions and is more tolerant of missing values and outliers. Due to the Law of Large Numbers, RF does not overfit. Injecting the correct randomness makes them accurate classifiers and regressors. RF algorithm can automatically distinguish the importance of each variable, give out the dependence between variables, and easily give explanations in combination with professional knowledge (Breiman, 2001). The RF method has been widely used for classification and regression in remote sensing applications (Ke et al., 2016; Park et al., 2016, 2018; Richardson et al., 2017). RF has begun to be used for Ta estimation in recent years (Ho et al., 2014; Zhang et al., 2016; Yoo et al., 2018). Use the default model parameter settings of R and its contribution packages to develop and apply statistical models (R Core Development Team, 2008; Ho et al., 2014; Liaw and Wiener, 2015).

Variable Adjustment

The absence of complete collinearity between any two independent variables is one of the assumptions of multiple linear regression. Variance Inflation Factor (VIF) is an index used to judge whether there is collinearity. If there is no linear relationship between the independent variables, the VIF value is 1, and a deviation from 1 indicates a trend of collinearity. From the effect of the multicollinearity test, multicollinearity can be tolerated when VIF <10. The VIF value of a variable greater than 10 indicates that there may be estimation problems. If there is multicollinearity, we would simply delete the variable directly, or use a biased estimate for processing (Shabani and Norouzi, 2015; Williams and Ojuri, 2021).

The Ta peaks around the 200th day of each year, which is nonlinear with the increase of Julian Day, so assuming that there is a quadratic function relationship. Linearize the Julian Day and adjust it to $x7 = (J - 200)^2$ as the independent variable. The Ts and the Julian day are independent variables with a strong correlation. Therefore, adding the relationship of $x8 = x7 \times Ts$ as an interaction term to match the model, assuming that the slope of Ts depends on the value of Julian day.

Selecting appropriate variables can not only avoid overfitting but also increase the explanatory degree of the model. The idea of the optimal subset selection method is to model all the variable combinations, then select the model with the best result. The advantage of this method is that all possible combinations are tested, and the final choice must be the best result. However, as the number of candidate variables increases, the amount of calculation will increase exponentially. Therefore, this method is only suitable for situations with few independent variables.

Residual Sum of Squares (RSS), adjusted coefficients of determination (adjusted R^2), Mallows's Cp (Cp) value, and Bayesian Information Criterion (BIC) value are used to evaluate model statistics (Cristóbal et al., 2008). The closer Adjusted R^2 is to 1, and the other indicators are smaller, the better the model fits. The optimal model can be determined by comparing the indicators of each variable.

Validation Data and Indicators

The verification of Ts used the infrared sensor (SI-111) observation data in the uniform and widespread farmland from 2014 to 2018. Ta and wind speed data records for 2014–2019 come from the daily data set of surface climate data, obtaining from the China Meteorological Data Service Centre (http://data.cma.cn/). Use Ts and other independent variables from 2014 to 2017 as a training set to build the MLR model and use data from 2018 to 2019 to verify the accuracy. The RF model randomly extracts 70% of all the data as a training set, using the remaining data to validate the resulting model. Such a verification method can explore whether the model constructed by the data from the past years is also applicable to the future years, which achieve the expansion of the time scale.

A set of statistical parameters were calculated to assess the accuracy of the predicted air temperature, including coefficients of determination (R^2), root mean square error (RMSE), and model efficiency (MEF). Values of R^2, RMSE, MEF and can be estimated using the following equations:

$$R^2 = \frac{\sum_{i=1}^{n}\left[(y_i - \bar{y})(O_i - \bar{O})\right]}{\sum_{i=1}^{n}\left[(y_i - \bar{y})^2\right]\sum_{i=1}^{n}\left[(O_i - \bar{O})^2\right]} \quad (8)$$

$$RMSE = \left[\sum_{i=1}^{n}\frac{(O_i - y_i)^2}{n}\right]^{\frac{1}{2}} \quad (9)$$

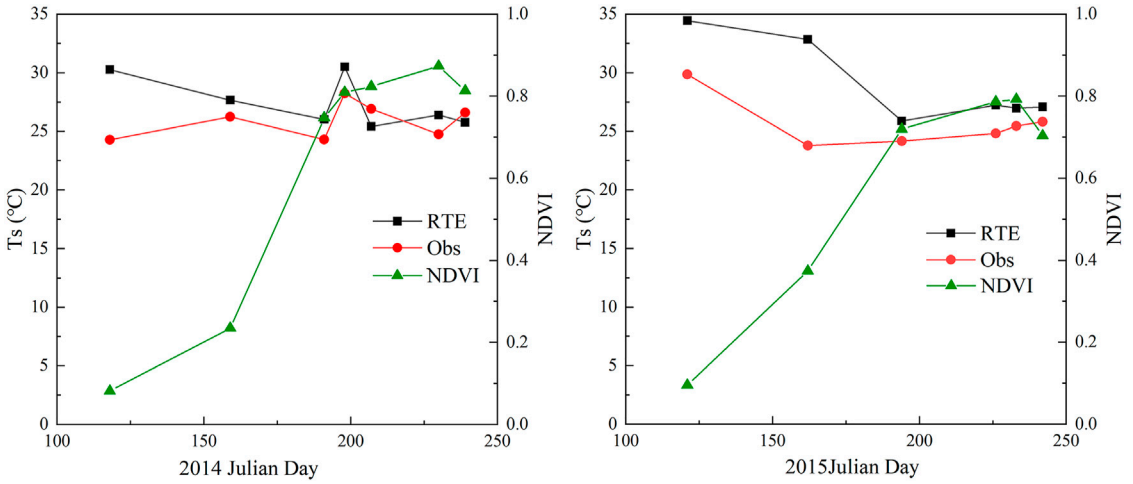


FIGURE 3 | Surface temperature (Ts °C) computed from the Radiative Transfer Equation (RTE) method and observed (Obs) from 2015 and 2014. As reference, the normalized difference vegetation index (NDVI) is included from Julian Day 100 to 250.

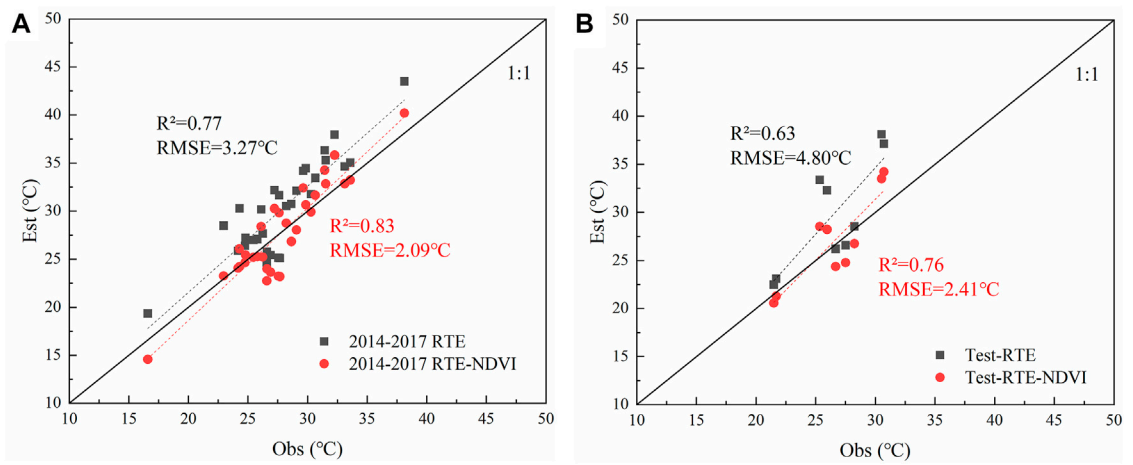


FIGURE 4 | Comparison between observed (Obs) and estimated (Est) surface temperature for 2014-2017 **(A)**, 2018 **(B)**. RTE and NDVI-RTE corresponds to the estimated values from the original equation and RTE equation adjusted using the normalized difference vegetation index (NDVI), respectively. Also, coefficients of determination (R2) and root mean square error (RMSE °C) are indicated in the figure.

$$\mathrm{MEF} = 1 - \left(\sum_{i=1}^{n} \frac{(O_i - y_i)^2}{\sum_{i=1}^{n} (O_i - \bar{O})^2} \right) \tag{10}$$

where n is the number of records in validation data sets used in this study, y_i is the estimated variable, and O_i is the observed variable, $\bar{y}$ is the mean value of the predicted for all validation sites, $\bar{O}$ is the mean value of the observed variable for all validation sites.

R^2 always calculated for a significance level of 0.005, was used as a measure of correlation and proportion of observed variability accounted by the model (Benali et al., 2012). RMSE is used to quantify the error (Willmott and Matsuura, 2005). RMSE is an indicator that shows the mean and spatial variance and is used to measure the quadratic error at a single level, which is also particularly sensitive to outliers (Janssen and Heuberger, 1995). The value of MEF is in the range between −1.0 and 1.0. If the performance of the estimation method is poor, the value will be lower (Zheng et al., 2013; Yang et al., 2017; Wang and Lu, 2018). Because MEF integrates correlation and error measurement, it is a robust statistical indicator for model consistency evaluation and reflects the adjustment of the 1:1 line, so it is used to measure the predictive ability of the model (Nash and Sutcliffe, 1970). The above parameters compared the

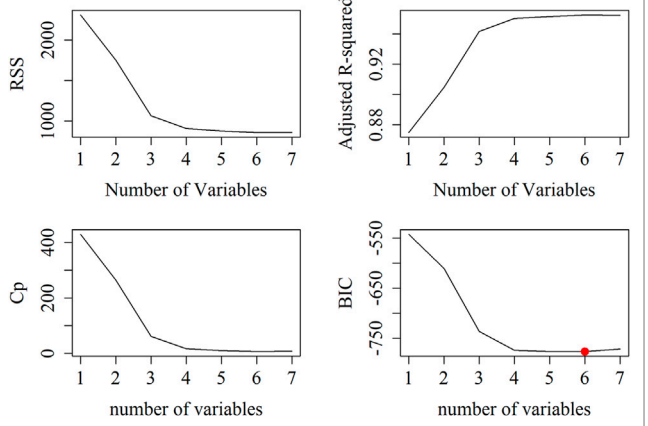


FIGURE 5 | Effect of number of variables in the air temperature model on Residual Sum of Squares (RSS), adjusted coefficients of determination (adjusted R^2), Mallows's Cp (Cp) value, and Bayesian Information Criterion (BIC).

observations and predictions values of each meteorological station and described the fitting performance of each model. The uncertainty of the predicted Ta in spatiotemporal scale was also calculated.

RESULTS

Surface Temperature Calibration

SI-111 collected Ts data for the 5 years from 2014 to 2018. Based on Landsat8, using RTE to invert the Ts, 40 points corresponding to the position of the ground instrument were extracted. The RTE method estimates Ts from 2014 to 2017, compared with observation data and its R^2 = 0.77, RMSE = 3.27°C. This study found that the RTE method overestimated the value of Ts. Especially when the surface vegetation coverage is low, the error will be more obvious (**Figure 3**). To resolve this uncertainty, a correction method using NDVI is proposed to improve the accuracy of the RTE method for inversion of Ts.

Correcting RTE with NDVI, the expression is:

$$Ts_{RTE-NDVI} = Ts_{RTE} - \frac{0.22}{NDVI} - 1.5 \tag{11}$$

where NDVI is Normalized Difference Vegetation Index; $Ts_{RTE-NDVI}$ is the surface temperature calculated using the Radiation Transfer Equation corrected by NDVI (°C); Ts_{RTE} is the surface temperature calculated using the original Radiation Transfer Equation (°C).

The 31 measured data from 2014 to 2017 were used as the target to revise the RTE (**Figure 4A**), and the 2018 data were used as verification (**Figure 4B**). After correction using NDVI of

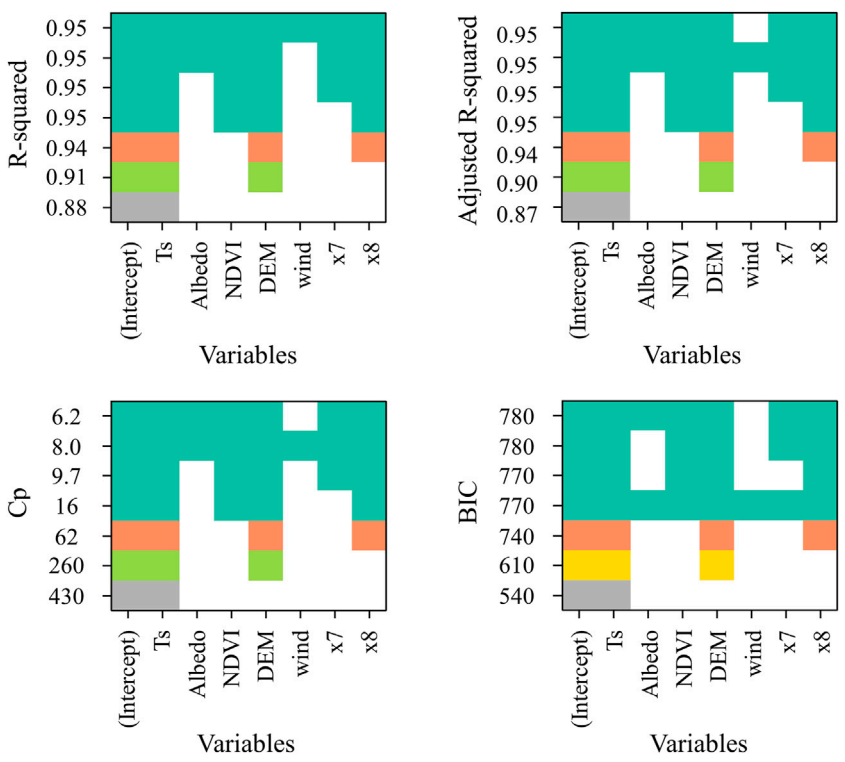


FIGURE 6 | Combination of different variables when the value of coefficients of determination (R^2), adjusted coefficients of determination (adjusted R^2), Mallows's Cp (Cp) value, and Bayesian Information Criterion (BIC) value were stabilized. The variables included surface temperature (Ts), normalized vegetation index (NDVI), surface albedo (Albedo), elevation (DEM), wind speed (wind), $x7 = (J - 200)^2$ and $x8 = x7 \times Ts$. The highest color in the figure indicates that the statistical indicators are gradually stable, and the variable combination is optimal.

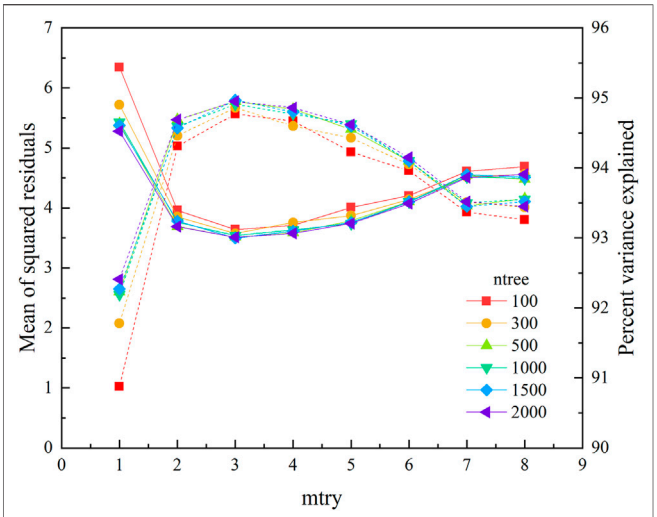


FIGURE 7 | Mean of squared residuals and variation in percentage variance explained by the random forest model with different number of trees to grow (ntree) and number of variables to be used at each node (mtry). Dotted lines represent percentage variance explained; solid lines represent Mean of squared residuals.

2014–2017, compared with the verified data, the accuracy of the model is R^2 = 0.83, RMSE = 2.09°C. And after the verification of the data in 2018, it is confirmed that **Eq. 11** has improved the accuracy of RTE (before: R^2 = 0.63, RMSE = 4.80°C; After: R^2 = 0.76, RMSE = 2.41°C). The fitting line is close to 1:1, indicating that the estimation accuracy of the Ts is significantly improved.

Variable Importance and Selection

Seven variables were tested for multicollinearity, and there was no variable with VIF greater than 10, indicating that the regression model can be established. As shown in **Figure 5**, RSS monotonically decreases with the increase of the number of independent variables, which cannot be directly judged. Adjusted R^2 and Cp values tended to be the largest when the number of variables was 6 and 7. The adjusted R^2 for Ta models ranged from 0.87 to 0.95, which eventually fixed at 0.95. Further addition of the independent variable did not improve the adjusted R^2, indicating that 6 independent variables were considered optimal. The change of BIC with the number of variables also proves that the model with six variables has the highest accuracy.

The statistical analysis indicated that Ts, Albedo, NDVI, DEM, x7, and x8 were significant independent variables for estimating Ta. As shown in **Figure 6**, R^2 was increased with the increase of the independent variables, so optimal combination of independent variables was not proposed in this study. However, in the absence of the wind speed, the adjusted R^2 reached a maximum while Cp and BIC arrived at a minimum, indicating that wind speed does not improve the accuracy of the Ta model. Except that the wind speed does not affect the model construction, the addition of other variables can improve the fitting accuracy of the model. The date of Ta estimation is almost always sunny when the average wind speed is slightly different in space. Therefore, there is no significant correlation between wind speed and Ta variation. Moreover, the average wind speed data is interpolated from the data obtained from meteorological stations, which may be inaccurate at the regional scale. To sum up, it is sufficient to use remote sensing data, elevation, and Julian days as the independent variables for the model.

After determining the independent variables, the formula of the constructed multiple linear regression equation is as follows:

$$Ta = b_0 + b_1 Ts + b_2 \alpha + b_3 NDVI + b_4 H + b_5 x7 + b_6 x8 \quad (12)$$

where Ta is the near-surface air temperature; Ts is the surface temperature; a is the surface albedo; NDVI is Normalized Difference Vegetation Index; H is the altitude;

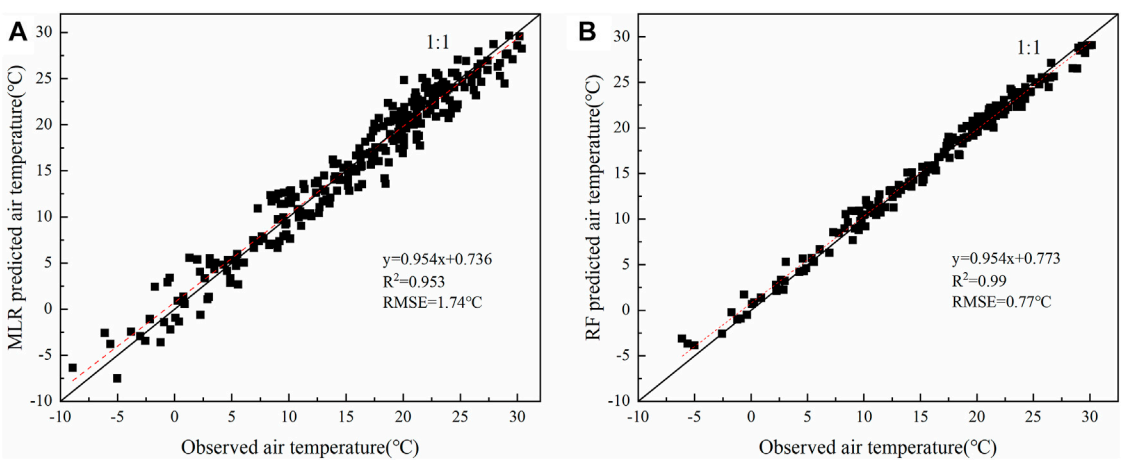


FIGURE 8 | Scatter plot between predicted and in-situ near-surface air temperatures (Ta °C) based on multiple linear regression (MLR) model (**Figure 8A**) and random forest (RF) algorithm (**Figure 8b**). The surface temperature (Ts °C) in the models is calculated by radiation transfer equation (RTE). Coefficients of determination (R^2) and root mean square error (RMSE °C) were used to evaluate the accuracy of Ta estimation from the training set.

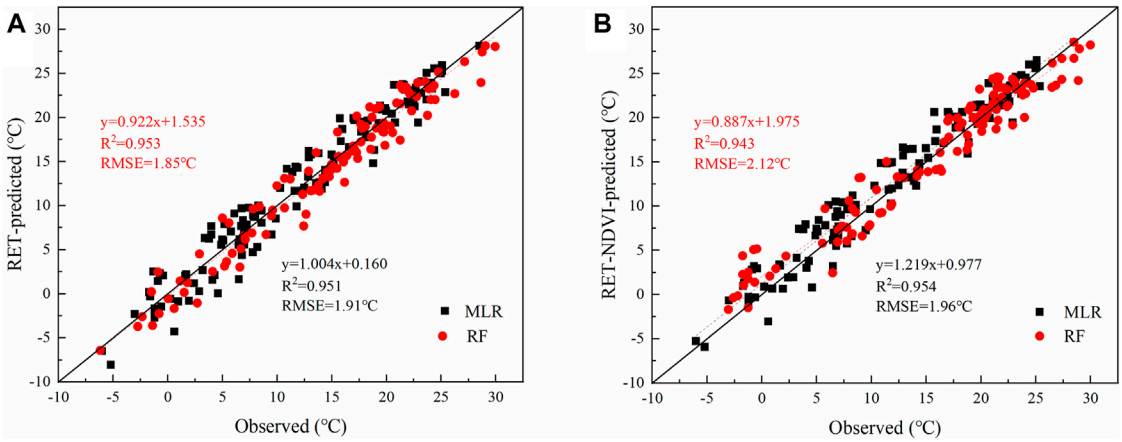


FIGURE 9 | The Ts involved in the Ta estimation were calculated through the radiation transfer equation (RTE, **(A)**) and the NDVI correction (NDVI-RTE, **(B)**). The coefficient of determination (R^2) and root mean square error (RMSE °C) were used to evaluate the accuracy of estimation Ta on the validation set for the two models and the two Ts data.

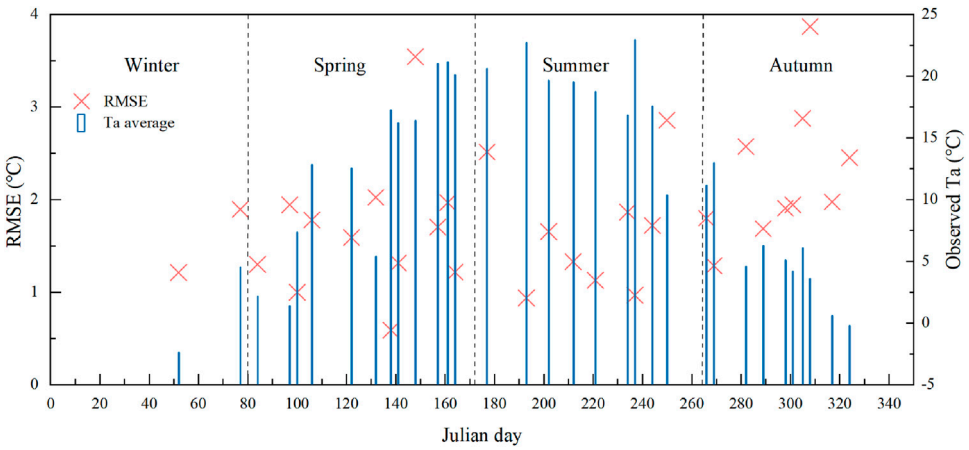


FIGURE 10 | Relations between the temporal distribution of observed near-surface air temperature (Ta °C) and model performance, using root mean square error (RMSE °C) to represent the accuracy in the estimation of air temperature (Ta).

$x7 = (J - 200)^2$; $x8 = x7 \times Ts$; J is Julian Day (the number of the day of the year); b_0~b_6 are undetermined coefficients.

The random forest algorithm cannot provide an estimation form similar to MLR because it cannot be parameterized. Two parameters are critical to the operation of the random forest model: the number of trees to grow (ntree) and the number of variables to be used at each node (mtry). These two parameters are selected based on the percentage of model interpretation and mean of squared residuals. The mtry value was tested from 1 to 8, and the ntree value was tested using the following 6 values: 100, 300, 500, 1,000, 1,500, and 2,000.

Figure 7 shows the changes in Mean of squared residuals and Percent variance explained after running a random forest with different ntree and mtry values. The highest Percent variance explained and the lowest Mean of squared residuals were obtained when mtry = 3 and ntree = 1,500 or 2,000. Since the higher the value of ntree, the more cost and time for calculation, so the value of ntree is set to 1,500, and the value of mtry is set to 3 to run the random forest model.

The amount of Ta-related data from 2014 to 2019 is 397. Randomly select 70% of the total amount as the training set to determine the variables and parameters, and the remaining data as the testing set. **Figure 8** is a scatter plot of the simulated and observed values of the two models constructed from the training set data. The R^2 and RMSE of the MLR model were 0.953 and 1.74°C, respectively. Most of the points of the RF are concentrated and close to the 1:1 line. The comparison between the results of the RF model and the MLR model shows that the random forest has higher estimation accuracy.

Model Validation

To understand the universality of the built model, the MLR model was verified using the 2018–2019 dataset; the RF algorithm was verified using the reserved test dataset. At the same time, it is

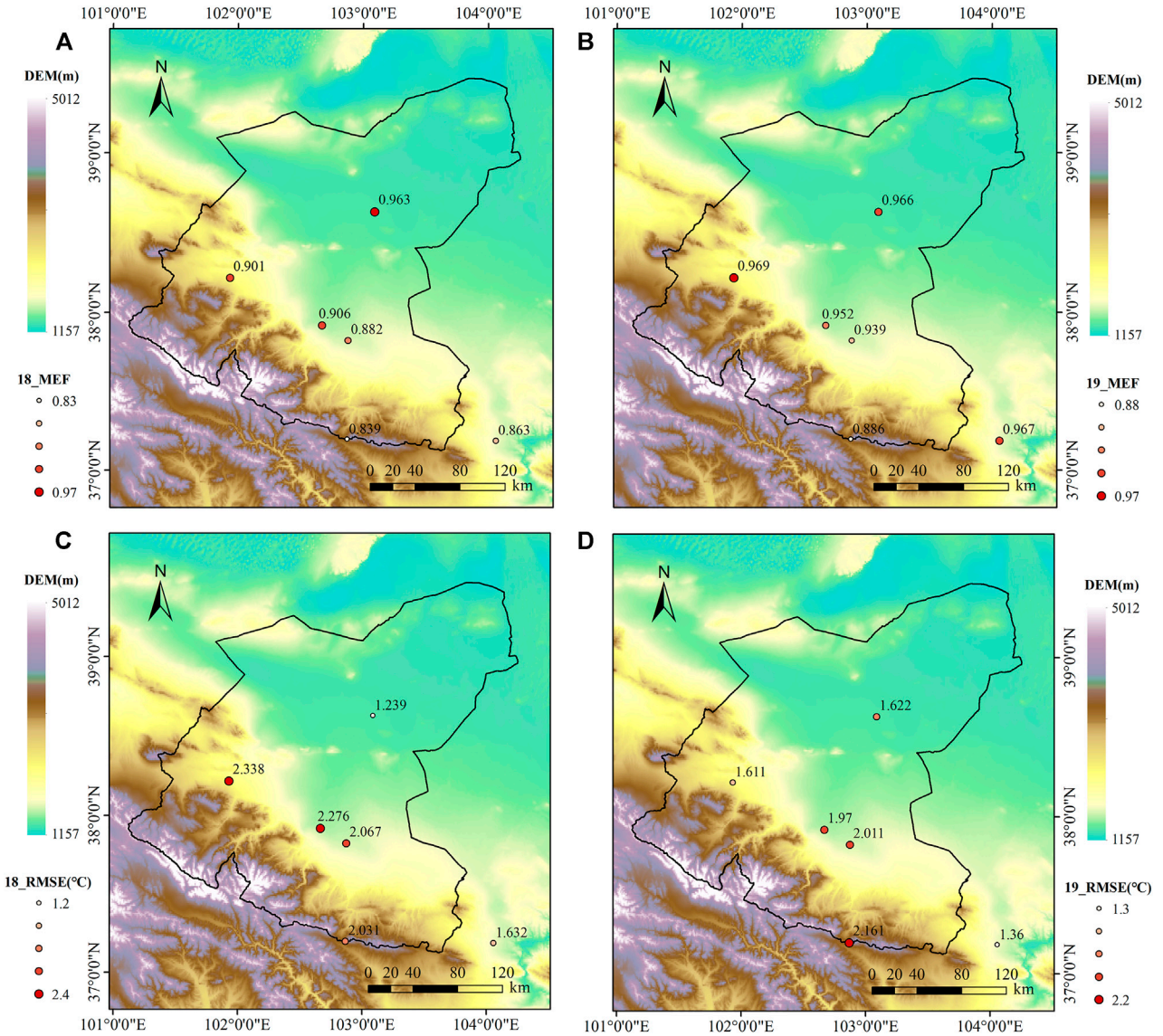


FIGURE 11 | Spatial distribution of the optimal models, model efficiency (MEF): **(A)** 2018, **(B)** 2019, root mean square error (RMSE °C): **(C)** 2018, **(D)** 2019 for all the meteorological stations.

explored whether the improvement of the estimation accuracy of Ts will affect the estimation accuracy of Ta.

The prediction result of Ta estimated through the validation data set (**Figure 9**) cannot reach the accuracy of the initial training set. When Ta is estimated from the validation data set of the RF, the prediction could not achieve the accuracy of the initial training set, respectively. When using NDVI to correct RTE, the R^2 and RMSE of the RF model was 0.943 and 2.12°C, respectively, indicating that the estimation accuracy of using RTE-NDVI is reduced. In contrast, the simulation results of MLR on Ta are relatively stable. The two calculation methods of Ts have little effect on MLR. Therefore, the accuracy of Ts inversion will not affect the accuracy of the two methods for estimating Ta. The estimation results of all models show good performance, and the fitting line is close to 1:1.

Spatial and Temporal Performance

Through the analysis of the average daily Ta and error estimated by MLR for 2018–2019, it was found that the temperature and error on the time scale have spatial variability. As shown in **Figure 10**, Ta had a clear trend with Julian day and season, and the estimation accuracy had an apparent seasonal trend. Sometimes, the meteorological station is covered by clouds and cannot participate in the calculation or verification, especially at the high-altitude Wushaoling station. The amount of remote sensing image data in winter is insufficient. Although the error is small, it cannot be fully verified. There was not much difference in the accuracy between spring and summer. The RMSE values in spring and summer were both 1.66°C, which had no difference in the estimation accuracy. With the increase of temperature, the error value was relatively stable and had no apparent change trend. In autumn, the RMSE was 2.35°C. At this

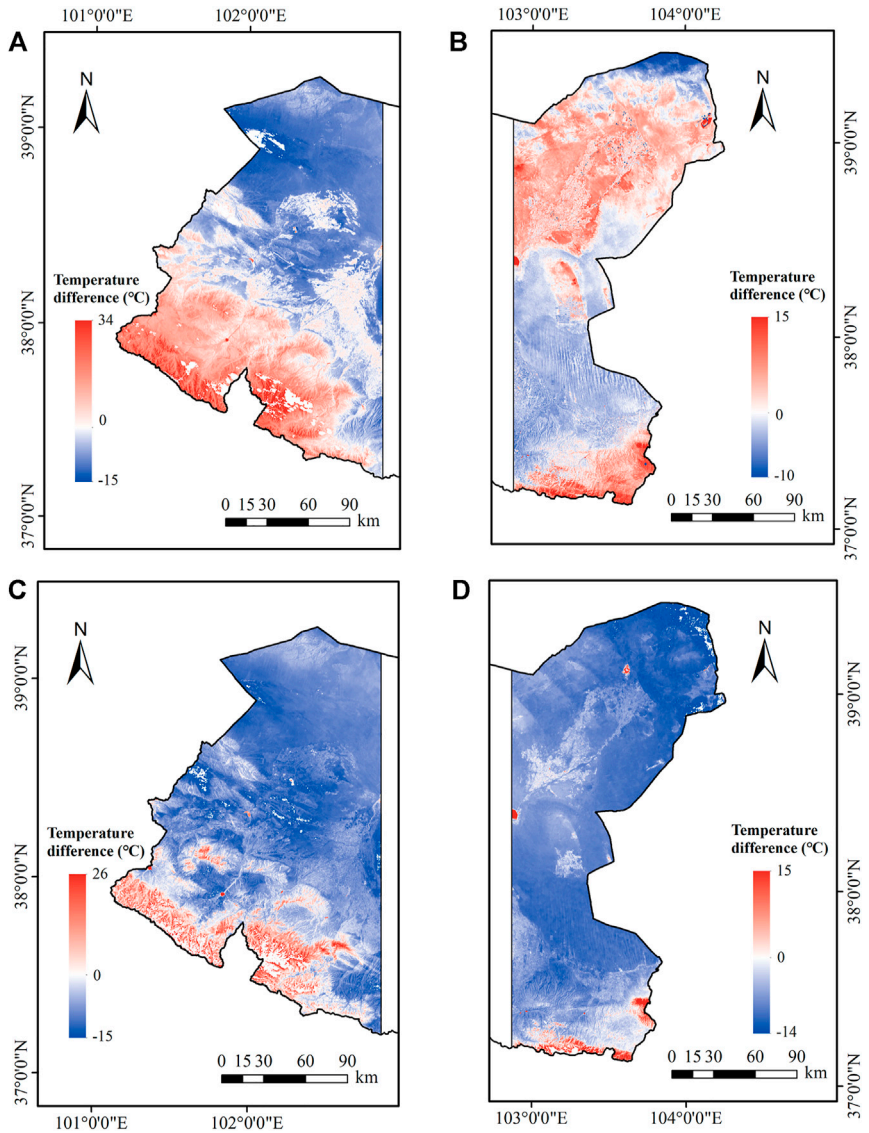


FIGURE 12 | Difference between interpolation method and multiple linear regression model estimation result **(A-D)**: (July 31 and August 22, 2018, and May 21 and May 28, 2019).

time, the study area is in the rainy season, and the estimate of Ta will be affected by the increase in rainfall.

In 2018 and 2019, the MEF value of each station was between 0.84 and 0.97, and the overall MEF was 0.946, indicating that MLR has good performance. Analyze the error distribution of Ta estimated by the MLR model at each site. **Figure 9** shows that Ta is more accurate at locations far from the mountainous area.

The MEF of the Minqin station for 2 years was close to 1, and the degree of the fitting was relatively high. The MLR model behaved differently in different altitude ranges. As the altitude increases, the performance of the MLR model gradually weakens, which may be caused by the complexity of the terrain. The RMSE has the maximum value at the highest altitude site (Wu Shaolin) (**Figures 11C,D**). The average RMSE of Minqin and Jintai Station are below 1.5°C. The accuracy of these two stations is better than that of other stations, possibly because the terrain and environmental conditions of the stations are not complicated. The locations of these two stations are far away from the Qilian Mountains, which are as high as 5,000 m above sea level. The estimation model had the highest relevance and accuracy at the Minqin station, which is located on the north side of the basin, and the surrounding terrain is flat and less affected by high mountains. Although Jintai Station is close to the east of Qilian Mountain, which is far from the mountain range, the estimation accuracy of Ta is also very high.

Because the data of the two adjacent columns of the Landsat satellite are of different periods, remote sensing images of the entire study area cannot be obtained on the same day, so the basin can only be divided into two parts for comparison. Chosen four images with fewer clouds, then calculated the difference

between the spatial interpolation result and the MLR model estimate. Inverse Distance Weighted (IDW) assumes that the influence of variables on the surrounding area decreases as the distance to the sample location increases, which was used to interpolate Ta data from the meteorological station. The error of Ta distribution obtained by IDW in the area is related to the underlying surface. Using meteorological and remote sensing data from July 31, August 22, 2018, May 21, May 28, 2019, Ta spatial interpolation and estimation maps were obtained and compared.

It iswidely believed that Ta is related to altitude. Temperature decreased with the increase of altitude in most cases. In high altitude areas, where no meteorological station provides temperature data, the Ta of the regions was determined by data from nearby stations. Therefore, Ta in high altitude areas was often overestimated by interpolation (the red area below **Figures 12A,C**). Simultaneously, there are also large desert areas in the study area (the blue area in **Figure 12**). Since the temperature in the desert during the day will be much higher than in other zones, the interpolation method can underestimate the temperature by up to 15°C. The Ta of farmland is overestimated during spatial interpolation (the part of the red area above d in **Figure 12B**), which is also related to the location of meteorological stations. Most of the station is established in cities or border areas. The vegetation coverage of farmland is very high during the growth period of crops, which will affect its Ta. Therefore, when using interpolation methods such as IDW, more meteorological stations are needed to ensure the accuracy of the interpolation results, and at the same time, should pay attention to the unreliability of interpolation methods in mountains or deserts.

DISCUSSION

Model Selection

In areas with vegetation cover, the temperature value retrieved from remote sensing data is often a mixed value of Ts and canopy temperature (Tc). The accurate estimation of Ts still needs to be analyzed according to the vegetation situation and terrain (Zhang and Li, 2018). Comparing the Ts calculated by RTE with the measured values, it was found that RTE overestimated the Ts, especially when the vegetation coverage was low, consistent with the conclusions published by USGS in 2015 (Barsi et al., 2014; Xu and Huang, 2016). However, due to the characteristics of fewer parameters and suitable for any thermal infrared band, the improved accuracy can also be widely used. Ts is considered to be the most important independent variable the many estimated Ta models (Park, 2011; Zhang et al., 2016; Song and Wu, 2018). In this study, the RMSE of Ts directly calculated by RTE is 3–5°C, and it is 2–3°C after correction. The correction equation is proposed based on the local estimation results, and its universality still needs to be verified. The error is related to the surface emissivity or atmospheric profile parameters (Li et al., 2013; Sekertekin and Bonafoni, 2020).

The variables of this study to be selected are closely related to Ta in many studies (Cristóbal et al., 2008; Gholamnia et al., 2017; Shen et al., 2020). Although the correlation between time parameters and Ta was not strong, the adjusted x7 had a significant influence on model accuracy indicating that the adjustment using the Julian Day was useful. Studies have suggested that wind speed is significant, especially when using energy balance methods, which require wind speed to calculate aerodynamic resistance (Hou et al., 2013; Zhang et al., 2015). Through indicators comparison, it is found that when using a statistical method to build the model, the average wind speed is not helpful to improve Ta estimation accuracy (Stisen et al., 2007).

The selected variables participate in MLR and RF to estimate Ta, using the available satellite dataset. The RMSE of the two methods were both lower than 2.0°C, which indicated that they were both suitable for northwest China. Previously, Ho et al. (2014) Using the RF algorithm to estimate the maximum Ta in Vancouver, Canada, and the RMSE was 2.3°C. The optimal MLR model established by Yang et al. (2017) estimates the average Ta and the RMSE is 3.6°C. Therefore, the methods of estimating Ta in this study had relatively high accuracy. However, due to the 16-days revisit cycle of the Landsat satellite and the low temporal resolution of the images, there are limitations in monitoring daily Ta (Yoo et al., 2018).

When performing regression, the random forest cannot make predictions that exceed the range of the training dataset, which may lead to overfitting when modeling some data with specific noise. For many statistical modelers, the random forest is like a black box (Zheng et al., 2019), almost unable to control the internal operation of the model, can only try between different parameters and random allocation. The simulation results of the random forest algorithm for the existing data are very good, but when it is used in the prediction and estimation, its accuracy may drop suddenly. It is difficult to improve it because it cannot control the internal operation of the model. For this study, multiple linear regression is still recommended. The simulation accuracy of the MLR model is similar to that of validation, which is conducive to the evaluation of subsequent prediction results. From the comparison results of the models, the MLR and RF models that without improving the accuracy of Ts performed better, and the actual operation was simpler. The reason for this phenomenon may be that the NDVI involved in the Ts correction is also one of the independent variables of the Ta estimation model, which makes the internal adjustment of the model during operation can ignore the errors of Ts.

Spatial and Temporal Uncertainties of Model

The seasonal variation of the average temperature and the distribution of the station will affect the estimation accuracy of Ta (Holden et al., 2011; Chen et al., 2015). From the time point of view, the RMSE in autumn reaches 2.24°C, which is 0.57°C higher than that in spring and summer. This is consistent with the results of Golkar et al. (2018), which believe that the estimation of Ta in spring and summer has higher accuracy, while there are more uncertainties in autumn. Therefore, the estimation model of daily average temperature is more suitable for spring and summer days. Yang et al. (2017) integrated various statistical indicators and believed that each model performed better in spring, and the estimation accuracy decreased due to the influence of rainfall and cloudy weather. The study area is located in the inland of Northwest China, and rainfall is mainly concentrated in autumn. Benali et al. (2012) believe that the cloud cover of remote sensing images is inversely proportional to the model performance, and higher cloud cover harms model

performance. That is, because cloud cover could reduce the accuracy of the Ts, which affects the accuracy of estimated values of daily average temperature. Therefore, the model can be optimized to the season or month scale, and the process of climate and environmental impacts can be added to adapt to the impact of climate change and cloud cover.

The spatial distribution map of RMSE shows that the model in mountainous and plateau areas generally performs worse than in plain areas. This conclusion is consistent with the research results of Yang et al. (2017). Due to the influence of terrain differences, mountainous terrain has a broader spatial variation range than flat terrain, so mountain temperature changes are more complicated. On relatively flat terrain and vegetation, horizontal uniformity leads to stable atmospheric conditions (Blandford et al., 2008; Lin et al., 2016). Ta raster data is usually limited by site coverage, especially in mountainous areas where the density and elevation distribution of meteorological observations vary greatly (Holden et al., 2011). Shen et al. (2020) believe that in the case of large-scale use of MLR and RF, the accuracy of the estimation results is poor due to the scarcity of sites in Northwest China that can participate in model training. When MLR is estimated in the Shiyang River Basin, the MEF of the highest altitude station (3000 m) is higher than 0.83, and the RMSE = 2.10°C, the model also performs well.

The method proposed by the research can obtain high-precision Ta estimation results when using the Landsat 8 data set, and improve the spatial resolution. However, the actual situation of only six meteorological stations limits the accuracy of Ta inversion in this area. We hope that more reliable verification data can be obtained in future studies.

CONCLUSION

The purpose of this study is to estimate Ts and Ta in the Shiyang River Basin. Perform parameter inversion based on Landsat8 images, use RTE to calculate and correct Ts. Using Ts, NDVI, elevation, and other parameters as variables to construct MLR and RF models for estimating Ta. The results show that: For MLR and RF, after calibrating Ts and participating in the estimation of Ta, both methods have accurate estimation results; The accuracy of RF training results is better than MLR, but the test set results of the two models are not significantly different; Although the topography of the study area is complex and the land cover conditions are different, the MLR model applies to the study area. In addition, compared with the MLR model, the comparison found that the interpolation method will underestimate the temperature in areas with low vegetation coverage like deserts, while the opposite is in the mountains and farmland areas.

This study can be used to accurately understand the distribution characteristics and changing trends of Ts and Ta in the study area. Further research can optimize the model in time to a month to adapt the climate change; make up the deficiency of Landsat8 and MODIS data by fusing remote sensing data, then obtaining Ta data with high temporal and spatial resolution.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

YL and SW conceived the idea. YL wrote the manuscript, SO, FT, and SW revised and improved the paper. The validation data of the model is provided by SL. All authors reviewed the manuscript.

FUNDING

This study was supported by the International and regional cooperation and exchange projects of the National Natural Science Foundation of China (51961125205), and the Chilean government through National Agency for Research and Development (ANID)/PCI (NSFC190013).

ACKNOWLEDGMENTS

The authors would like to thank the NASA team for making the remote sensing data freely available, the GEE platform for processing remote sensing data, and the National Meteorological Information Center of China Meteorological Administration for archiving the observed climate data (http://data.cma.cn).

REFERENCES

Agha, S. R., and Alnahhal, M. J. (2012). Neural Network and Multiple Linear Regression to Predict School Children Dimensions for Ergonomic School Furniture Design. *Appl. Ergon.* 43, 979–984. doi:10.1016/j.apergo.2012.01.007

Akan, R., Keskin, S. N., and Uzundurukan, S. (2015). Multiple Regression Model for the Prediction of Unconfined Compressive Strength of Jet Grout Columns. *Proced. Earth Planet. Sci.* 15, 299–303. doi:10.1016/j.proeps.2015.08.072

Alkama, R., and Cescatti, A. (2016). Biophysical Climate Impacts of Recent Changes in Global forest Cover. *Science* 351, 600–604. doi:10.1126/science.aac8083

Anderson, M., Norman, J., Kustas, W., Houborg, R., Starks, P., and Agam, N. (2008). A thermal-based Remote Sensing Technique for Routine Mapping of Land-Surface Carbon, Water and Energy Fluxes from Field to Regional Scales. *Remote Sensing Environ.* 112, 4227–4241. doi:10.1016/j.rse.2008.07.009

Bai, Y., Wu, L., Qin, K., Zhang, Y., Shen, Y., and Zhou, Y. (2016). A Geographically and Temporally Weighted Regression Model for Ground-Level PM2.5 Estimation from Satellite-Derived 500 M Resolution AOD. *Remote Sensing* 8, 262. doi:10.3390/rs8030262

Barsi, J., Schott, J., Hook, S., Raqueno, N., Markham, B., and Radocinski, R. (2014). Landsat-8 thermal Infrared Sensor (TIRS) Vicarious Radiometric Calibration. *Remote Sensing* 6, 11607–11626. doi:10.3390/rs61111607

Bathiany, S., Dakos, V., Scheffer, M., and Lenton, T. M. (2018). Climate Models Predict Increasing Temperature Variability in Poor Countries. *Sci. Adv.* 4, 1–11. doi:10.1126/sciadv.aar5809

Becker, W. R., Ló, T. B., Johann, J. A., and Mercante, E. (2021). Statistical Features for Land Use and Land Cover Classification in Google Earth Engine.

Remote Sensing Appl. Soc. Environ. 21, 100459. doi:10.1016/j.rsase.2020.100459

Benali, A., Carvalho, A. C., Nunes, J. P., Carvalhais, N., and Santos, A. (2012). Estimating Air Surface Temperature in Portugal Using MODIS LST Data. *Remote Sensing Environ.* 124, 108–121. doi:10.1016/j.rse.2012.04.024

Blandford, T. R., Humes, K. S., Harshburger, B. J., Moore, B. C., Walden, V. P., and Ye, H. (2008). Seasonal and Synoptic Variations in Near-Surface Air Temperature Lapse Rates in a Mountainous basin. *J. Appl. Meteorol. Climatol.* 47, 249–261. doi:10.1175/2007JAMC1565.1

Breiman, L. (2001). Random Forests. *Mach. Learn.* 45, 5–32. doi:10.1023/A:1010933404324

Chen, F., Liu, Y., Liu, Q., and Qin, F. (2015). A Statistical Method Based on Remote Sensing for the Estimation of Air Temperature in China. *Int. J. Climatol* 35, 2131–2143. doi:10.1002/joc.4113

Cresswell, M. P., Morse, A. P., Thomson, M. C., and Connor, S. J. (1999). Estimating Surface Air Temperatures, from Meteosat Land Surface Temperatures, Using an Empirical Solar Zenith Angle Model. *Int. J. Remote Sensing* 20, 1125–1132. doi:10.1080/014311699212885

Cristóbal, J., Ninyerola, M., and Pons, X. (2008). Modeling Air Temperature through a Combination of Remote Sensing and GIS Data. *J. Geophys. Res.* 113, 1–13. doi:10.1029/2007JD009318

Czajkowski, K. P., Mulhern, T., Goward, S. N., Cihlar, J., Dubayah, R. O., and Prince, S. D. (1997). Biospheric Environmental Monitoring at BOREAS with AVHRR Observations. *J. Geophys. Res.* 102, 29651–29662. doi:10.1029/97jd01327

Emamifar, S., Rahimikhoob, A., and Noroozi, A. A. (2013). Daily Mean Air Temperature Estimation from MODIS Land Surface Temperature Products Based on M5 Model Tree. *Int. J. Climatol.* 33, 3174–3181. doi:10.1002/joc.3655

Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., et al. (2007). The Shuttle Radar Topography Mission. *Rev. Geophys.* 45, RG2004. doi:10.1029/2005RG000183

Folland, C. K., Boucher, O., Colman, A., and Parker, D. E. (2018). Causes of Irregularities in Trends of Global Mean Surface Temperature since the Late 19th century. *Sci. Adv.* 4. doi:10.1126/sciadv.aao5297

Genuer, R., Poggi, J.-M., Tuleau-Malot, C., and Villa-Vialaneix, N. (2017). Random Forests for Big Data. *Big Data Res.* 9, 28–46. doi:10.1016/j.bdr.2017.07.003

Gholamnia, M., Darvishi Boloorani, A., Hamzeh, S., Kiavarz, M., and Kiavarz, M. (2017). Diurnal Air Temperature Modeling Based on the Land Surface Temperature. *Remote Sensing* 9, 915. doi:10.3390/rs9090915

Giacomino, A., Abollino, O., Malandrino, M., and Mentasti, E. (2011). The Role of Chemometrics in Single and Sequential Extraction Assays: A Review. Part II. Cluster Analysis, Multiple Linear Regression, Mixture Resolution, Experimental Design and Other Techniques. *Analytica Chim. Acta* 688, 122–139. doi:10.1016/j.aca.2010.12.028

Golkar, F., Sabziparvar, A. A., Khanbilvardi, R., Nazemosadat, M. J., Zand- Parsa, S., and Rezaei, Y. (2018). Estimation of Instantaneous Air Temperature Using Remote Sensing Data. *Int. J. Remote Sensing* 39, 258–275. doi:10.1080/01431161.2017.1382743

Goward, S. N., Waring, R. H., Dye, D. G., and Yang, J. (1994). Ecological Remote Sensing at OTTER: Satellite Macroscale Observations. *Ecol. Appl.* 4, 322–343. doi:10.2307/1941937

Hachem, S., Duguay, C. R., and Allard, M. (2012). Comparison of MODIS-Derived Land Surface Temperatures with Ground Surface and Air Temperature Measurements in Continuous Permafrost Terrain. *The Cryosphere* 6, 51–69. doi:10.5194/tc-6-51-2012

Ho, H. C., Knudby, A., Sirovyak, P., Xu, Y., Hodul, M., and Henderson, S. B. (2014). Mapping Maximum Urban Air Temperature on Hot Summer Days. *Remote Sensing Environ.* 154, 38–45. doi:10.1016/j.rse.2014.08.012

Holden, Z. A., Abatzoglou, J. T., Luce, C. H., and Baggett, L. S. (2011). Empirical Downscaling of Daily Minimum Air Temperature at Very fine Resolutions in Complex Terrain. *Agric. For. Meteorology* 151, 1066–1073. doi:10.1016/j.agrformet.2011.03.011

Hou, P., Chen, Y., Qiao, W., Cao, G., Jiang, W., and Li, J. (2013). Near-surface Air Temperature Retrieval from Satellite Images and Influence by Wetlands in Urban Region. *Theor. Appl. Climatol.* 111, 109–118. doi:10.1007/s00704-012-0629-7

Isaya Ndossi, M., and Avdan, U. (2016). Application of Open Source Coding Technologies in the Production of Land Surface Temperature (LST) Maps from Landsat: A PyQGIS Plugin. *Remote Sensing* 8, 413. doi:10.3390/rs8050413

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An Introduction to Statistical Learning: With Applications in R.* 1st ed. New YorkNew York: Springer.

Jang, J.-D., Viau, A. A., and Anctil, F. (2004). Neural Network Estimation of Air Temperatures from AVHRR Data. *Int. J. Remote Sensing* 25, 4541–4554. doi:10.1080/01431160310001657533

Janssen, P. H. M., and Heuberger, P. S. C. (1995). Calibration of Process-Oriented Models. *Ecol. Model.* 83, 55–66. doi:10.1016/0304-3800(95)00084-9

Jimenez-Munoz, J. C., Cristobal, J., Sobrino, J. A., Soria, G., Ninyerola, M., Pons, X., et al. (2009). Revision of the Single-Channel Algorithm for Land Surface Temperature Retrieval from Landsat thermal-infrared Data. *IEEE Trans. Geosci. Remote Sensing* 47, 339–349. doi:10.1109/TGRS.2008.2007125

Ke, Y., Im, J., Park, S., and Gong, H. (2016). Downscaling of MODIS One Kilometer Evapotranspiration Using Landsat-8 Data and Machine Learning Approaches. *Remote Sensing* 8, 215–226. doi:10.3390/rs8030215

Krüger, E., and Emmanuel, R. (2013). Accounting for Atmospheric Stability Conditions in Urban Heat Island Studies: The Case of Glasgow, UK. *Landscape Urban Plann.* 117, 112–121. doi:10.1016/j.landurbplan.2013.04.019

Li, L., and Zha, Y. (2018). Mapping Relative Humidity, Average and Extreme Temperature in Hot Summer over China. *Sci. Total Environ.* 615, 875–881. doi:10.1016/j.scitotenv.2017.10.022

Li, S., Quackenbush, L. J., and Im, J. (2019). Airborne Lidar Sampling Strategies to Enhance forest Aboveground Biomass Estimation from Landsat Imagery. *Remote Sensing* 11, 1906. doi:10.3390/rs11161906

Li, X., Zhou, Y., Asrar, G. R., and Zhu, Z. (2018). Developing a 1 Km Resolution Daily Air Temperature Dataset for Urban and Surrounding Areas in the Conterminous United States. *Remote Sensing Environ.* 215, 74–84. doi:10.1016/j.rse.2018.05.034

Li, Z.-L., Tang, B.-H., Wu, H., Ren, H., Yan, G., Wan, Z., et al. (2013). Satellite-derived Land Surface Temperature: Current Status and Perspectives. *Remote Sensing Environ.* 131, 14–37. doi:10.1016/j.rse.2012.12.008

Liang, S., Shuey, C. J., Russ, A. L., Fang, H., Chen, M., Walthall, C. L., et al. (2003). Narrowband to Broadband Conversions of Land Surface Albedo: II. Validation. *Remote Sensing Environ.* 84, 25–41. doi:10.1016/S0034-4257(02)00068-8

Liaw, A., and Wiener, M. (2015). *randomForest: Breiman and Cutler's Random Forests for Classification and Regression*.

Lin, X., Zhang, W., Huang, Y., Sun, W., Han, P., Yu, L., et al. (2016). Empirical Estimation of Near-Surface Air Temperature in China from MODIS LST Data by Considering Physiographic Features. *Remote Sensing* 8, 629. doi:10.3390/rs8080629

Ma, S., and Pu, Z. (2020). Research on Surface Temperature Inversion Algorithm Based on Landsat8 Data in Urumqi City. *Comput. Digit. Eng.* 10, 2316–2320. doi:10.3969/j.issn.1672-9722.2020.10.003

Meyer, H., Katurji, M., Appelhans, T., Müller, M., Nauss, T., Roudier, P., et al. (2016). Mapping Daily Air Temperature for Antarctica Based on MODIS LST. *Remote Sensing* 8, 732. doi:10.3390/rs8090732

Moser, G., De Martino, M., and Serpico, S. B. (2015). Estimation of Air Surface Temperature from Remote Sensing Images and Pixelwise Modeling of the Estimation Uncertainty through Support Vector Machines. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing* 8, 332–349. doi:10.1109/JSTARS.2014.2361862

Mostovoy, G. V., King, R. L., Reddy, K. R., Kakani, V. G., and Filippova, M. G. (2006). Statistical Estimation of Daily Maximum and Minimum Air Temperatures from MODIS LST Data over the State of Mississippi. *GIScience & Remote Sensing* 43, 78–110. doi:10.2747/1548-1603.43.1.78

Nash, J. E., and Sutcliffe, J. V. (1970). River Flow Forecasting through Conceptual Models Part III - the Ray Catchment at Grendon Underwood. *J. Hydrol.* 11, 109–128. doi:10.1016/0022-1694(70)90098-3

Nemani, R. R., and Running, S. W. (1989). Estimation of Regional Surface Resistance to Evapotranspiration from NDVI and Thermal-IR AVHRR Data. *J. Appl. Meteorol.* 28, 276–284. doi:10.1175/1520-0450(1989)028<0276

Nieto, H., Sandholt, I., Aguado, I., Chuvieco, E., and Stisen, S. (2011). Air Temperature Estimation with MSG-SEVIRI Data: Calibration and Validation of the TVX Algorithm for the Iberian Peninsula. *Remote Sensing Environ.* 115, 107–116. doi:10.1016/j.rse.2010.08.010

Noi, P., Degener, J., and Kappas, M. (2017). Comparison of Multiple Linear Regression, Cubist Regression, and Random forest Algorithms to Estimate Daily Air Surface Temperature from Dynamic Combinations of MODIS LST Data. *Remote Sensing* 9, 398. doi:10.3390/rs9050398

Orhan, O., Ekercin, S., and Dadaser-Celik, F. (2014). Use of Landsat Land Surface Temperature and Vegetation Indices for Monitoring Drought in the Salt Lake Basin Area, Turkey. *Scientific World J.* 2014, 1–11. doi:10.1155/2014/142939

Orhan, O., and Yakar, M. (2016). Investigating Land Surface Temperature Changes Using Landsat Data in Konya, Turkey. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* XLI-B8, 285–289. doi:10.5194/isprsarchives-XLI-B8-285-2016

Park, S., Im, J., Jang, E., and Rhee, J. (2016). Drought Assessment and Monitoring through Blending of Multi-Sensor Indices Using Machine Learning Approaches for Different Climate Regions. *Agric. For. Meteorology* 216, 157–169. doi:10.1016/j.agrformet.2015.10.011

Park, S. (2011). Integration of Satellite-Measured LST Data into Cokriging for Temperature Estimation on Tropical and Temperate Islands. *Int. J. Climatol.* 31, 1653–1664. doi:10.1002/joc.2185

Park, S., Seo, E., Kang, D., Im, J., and Lee, M.-I. (2018). Prediction of Drought on Pentad Scale Using Remote Sensing Data and MJO index through Random forest over East Asia. *Remote Sensing* 10, 1811–1818. doi:10.3390/rs10111811

Prince, S. D., Goetz, S. J., Dubayah, R. O., Czajkowski, K. P., and Thawley, M. (1998). Inference of Surface and Air Temperature, Atmospheric Precipitable Water and Vapor Pressure Deficit Using Advanced Very High-Resolution Radiometer Satellite Observations: Comparison with Field Observations. *J. Hydrol.* 212-213, 230–249. doi:10.1016/S0022-1694(98)00210-8

R Core Development Team (2008). *R: A Language and Environment for Statistical Computing*.

Raja Reddy, K., Hodges, H. F., and Mckinion, J. M. (1997). Crop Modeling and Applications: A Cotton Example. *Adv. Agron.* 59, 225–290. doi:10.1016/S0065-2113(08)60056-5

Richardson, H. J., Hill, D. J., Denesiuk, D. R., and Fraser, L. H. (2017). A Comparison of Geographic Datasets and Field Measurements to Model Soil Carbon Using Random Forests and Stepwise Regressions (British Columbia, Canada). *GIScience & Remote Sensing* 54, 573–591. doi:10.1080/15481603.2017.1302181

Riddering, J. P., and Queen, L. P. (2006). Estimating Near-Surface Air Temperature with NOAA AVHRR. *Can. J. Remote Sensing* 32, 33–43. doi:10.5589/m06-004

Ruiz-Álvarez, M., Alonso-Sarria, F., and Gomariz-Castillo, F. (2019). Interpolation of Instantaneous Air Temperature Using Geographical and MODIS Derived Variables with Machine Learning Techniques. *Ijgi* 8, 382. doi:10.3390/ijgi8090382

Sekertekin, A., and Bonafoni, S. (2020). Land Surface Temperature Retrieval from Landsat 5, 7, and 8 over Rural Areas: Assessment of Different Retrieval Algorithms and Emissivity Models and Toolbox Implementation. *Remote Sensing* 12, 294. doi:10.3390/rs12020294

Shamir, E., and Georgakakos, K. P. (2014). MODIS Land Surface Temperature as an index of Surface Air Temperature for Operational Snowpack Estimation. *Remote Sensing Environ.* 152, 83–98. doi:10.1016/j.rse.2014.06.001

Shen, H., Jiang, Y., Li, T., Cheng, Q., Zeng, C., and Zhang, L. (2020). Deep Learning-Based Air Temperature Mapping by Fusing Remote Sensing, Station, Simulation and Socioeconomic Data. *Remote Sensing Environ.* 240, 111692. doi:10.1016/j.rse.2020.111692

Shabani, A., and Norouzi, M. (2015). Predicting Cation Exchange Capacity by Artificial Neural Network and Multiple Linear Regression Using Terrain and Soil Characteristics. *Indian J. Sci. Technol.* 8. doi:10.17485/ijst/2015/v8i28/83328

Shi, L., Liu, P., Kloog, I., Lee, M., Kosheleva, A., and Schwartz, J. (2016). Estimating Daily Air Temperature across the Southeastern United States Using High-Resolution Satellite Data: A Statistical Modeling Study. *Environ. Res.* 146, 51–58. doi:10.1016/j.envres.2015.12.006

Sobrino, J. A., Jiménez-Muñoz, J. C., and Paolini, L. (2004). Land Surface Temperature Retrieval from LANDSAT TM 5. *Remote Sensing Environ.* 90, 434–440. doi:10.1016/j.rse.2004.02.003

Song, Y., and Wu, C. (2018). Examining Human Heat Stress with Remote Sensing Technology. *GIScience & Remote Sensing* 55, 19–37. doi:10.1080/15481603.2017.1354804

Stisen, S., Sandholt, I., Nørgaard, A., Fensholt, R., and Eklundh, L. (2007). Estimation of Diurnal Air Temperature Using MSG SEVIRI Data in West Africa. *Remote Sensing Environ.* 110, 262–274. doi:10.1016/j.rse.2007.02.025

Vancutsem, C., Ceccato, P., Dinku, T., and Connor, S. J. (2010). Evaluation of MODIS Land Surface Temperature Data to Estimate Air Temperature in Different Ecosystems over Africa. *Remote Sensing Environ.* 114, 449–465. doi:10.1016/j.rse.2009.10.002

Vázquez, D. P., Reyes, F. J. O., and Arboledas, L. A. (1997). A Comparative Study of Algorithms for Estimating Land Surface Temperature from AVHRR Data. *Remote Sens. Environ.* 62, 215–222. doi:10.1016/S0034-4257(97)00091-6

Vogt, J. V., Viau, A. A., and Paquet, F. (1997). Mapping Regional Air Temperature fields Using Satellite-Derived Surface Skin Temperatures. *Int. J. Climatol.* 17, 1559–1579. doi:10.1002/(sici)1097-0088

Wang, M., He, G., Zhang, Z., Wang, G., Zhang, Z., Cao, X., et al. (2017). Comparison of Spatial Interpolation and Regression Analysis Models for an Estimation of Monthly Near Surface Air Temperature in China. *Remote Sensing* 9, 1278. doi:10.3390/rs9121278

Wang, M. M., He, G. J., Zhang, Z. M., Zhang, Z. J., and Liu, X. G. (2018). Estimation of Monthly Near Surface Air Temperature Using Geographically Weighted Regression in China. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* XLII-3, 1747–1750. doi:10.5194/isprs-archives-XLII-3-1747-2018

Wang, W., and Lu, Y. (2018). Analysis of the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE) in Assessing Rounding Model. *IOP Conf. Ser. Mater. Sci. Eng.* 324, 012049. doi:10.1088/1757-899X/324/1/012049

Wang, Y., Zhou, J., Li, M., and Zhang, X. (2016). Validation of Landsat-8 TIRS LAND Surface Temperature Retrieved from Multiple Algorithms in an Extremely Arid Region. *Int. Geosci. Remote Sens. Symp.*, 6934–6937. doi:10.1109/IGARSS.2016.7730809

Williams, C. G., and Ojuri, O. O. (2021). Predictive Modelling of Soils' Hydraulic Conductivity Using Artificial Neural Network and Multiple Linear Regression. *SN Appl. Sci.* 3, 7. doi:10.1007/s42452-020-03974-7

Willmott, C., and Matsuura, K. (2005). Advantages of the Mean Absolute Error (MAE) over the Root Mean Square Error (RMSE) in Assessing Average Model Performance. *Clim. Res.* 30, 79–82. doi:10.3354/cr030079

Windahl, E., and Beurs, K. d. (2016). An Intercomparison of Landsat Land Surface Temperature Retrieval Methods under Variable Atmospheric Conditions Using *In Situ* Skin Temperature. *Int. J. Appl. Earth Observation Geoinformation* 51, 11–27. doi:10.1016/j.jag.2016.04.003

Xu, H. Q., and Huang, S. L. (2016). A Comparative Study on the Calibration Accuracy of Landsat 8 thermal Infrared Sensor Data. *Spectrosc. Spectr. Anal.*, 1941–1948. doi:10.3964/j.issn.1000-0593(2016)06-1941-08

Xu, H. Q. (2015). Retrieval of the Reflectance and Land Surface Temperature of the Newly-Launched Landsat 8 Satellite. *Chin. J. Geophys. Chin. Ed.* 58, 741–747. doi:10.6038/cij20150304

Xu, Y., Qin, Z., and Shen, Y. (2012). Study on the Estimation of Near-Surface Air Temperature from MODIS Data by Statistical Methods. *Int. J. Remote Sensing* 33, 7629–7643. doi:10.1080/01431161.2012.701351

Yang, Y., Cai, W., and Yang, J. (2017). Evaluation of MODIS Land Surface Temperature Data to Estimate Near-Surface Air Temperature in Northeast China. *Remote Sensing* 9, 410–419. doi:10.3390/rs9050410

Yoo, C., Im, J., Park, S., and Quackenbush, L. J. (2018). Estimation of Daily Maximum and Minimum Air Temperatures in Urban Landscapes Using MODIS Time Series Satellite Data. *ISPRS J. Photogrammetry Remote Sensing* 137, 149–162. doi:10.1016/j.isprsjprs.2018.01.018

Zakšek, K., and Schroedter-Homscheidt, M. (2009). Parameterization of Air Temperature in High Temporal and Spatial Resolution from a Combination of the SEVIRI and MODIS Instruments. *ISPRS J. Photogrammetry Remote Sensing* 64, 414–421. doi:10.1016/j.isprsjprs.2009.02.006

Zhang, H., Zhang, F., Ye, M., Che, T., and Zhang, G. (2016). Estimating Daily Air Temperatures over the Tibetan Plateau by Dynamically Integrating MODIS LST Data. *J. Geophys. Res. Atmos.* 121 (11), 11,425–11,441. doi:10.1002/2016JD025154

Zhang, R., Rong, Y., Tian, J., Su, H., Li, Z.-L., and Liu, S. (2015). A Remote Sensing Method for Estimating Surface Air Temperature and Surface Vapor Pressure on a Regional Scale. *Remote Sensing* 7, 6005–6025. doi:10.3390/rs70506005

Zhang, X., and Li, P. (2018). A Temperature and Vegetation Adjusted NTL Urban index for Urban Area Mapping and Analysis. *ISPRS J. Photogrammetry Remote Sensing* 135, 93–111. doi:10.1016/j.isprsjprs.2017.11.016

Zheng, L., Jinying, X. U., and Wang, X. (2019). Application of Random Forests Algorithm in Researches on Wetlands. *Wetl. Sci.* 01, 16–24. doi:10.13248/j.cnki.wetlandsci.2019.01.003

Zheng, X., Zhu, J., and Yan, Q. (2013). Monthly Air Temperatures over Northern China Estimated by Integrating MODIS Data with GIS Techniques. *J. Appl. Meteorol. Climatol.* 52, 1987–2000. doi:10.1175/JAMC-D-12-0264.1

Zhu, W., Lű, A., and Jia, S. (2013). Estimation of Daily Maximum and Minimum Air Temperature Using MODIS Land Surface Temperature Products. *Remote Sensing Environ.* 130, 62–73. doi:10.1016/j.rse.2012.10.034

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Environmental Science

ORIGINAL RESEARCH
published: 10 February 2022
doi: 10.3389/fenvs.2022.810902

Assessing Spatial and Temporal Distribution of Algal Blooms Using Gini Coefficient and Lorenz Asymmetry Coefficient

Ting Zhou[1], *Cheng Ni*[1], *Ming Zhang*[2] *and Ping Xia*[1]*

[1]School of Engineering, Anhui Agricultural University, Hefei, China, [2]School of Civil Engineering, Anhui Polytechnic University, Wuhu, China

Algal bloom in an inland lake is characterized by significant spatial and temporal dynamics. Accurate assessment of algal bloom distribution and dynamics is highly required for tracing the causes of and creating countermeasures for algal bloom. Satellite remote sensing provides a fast and efficient way to capture algal bloom distribution at a large scale, but it is difficult to directly derive accurate and quantitative assessment based on satellite images. In this study, the Gini coefficient and Lorenz asymmetry coefficient were introduced to examine the spatio-temporal algal bloom distribution of Chaohu Lake, the fifth largest inland lake in China. A total of 61 remote sensing images from three satellite sensors, Landsat, Gaofen, and Sentinel were selected to obtain algal bloom distributions. By dividing remote sensing images into 0.01°*0.01° grid cells, the normalized difference vegetation index (NDVI) for each grid cell was derived, forming a spatial and time series database for quantitative analysis. Two coefficients, Gini coefficient and Lorenz asymmetry coefficient, were used to evaluate the overall intensity, unevenness, and attribution of algal bloom in Chaohu Lake from 2011 to 2020. The Gini coefficient results show a large variety of algal bloom in the spatial and temporal scales of Chaohu lake. The lake edge and northwestern part had longer lasting and more severe algal bloom than the lake center, which was mainly due to nutrient import, especially from three northwestern tributaries that flow through the upstream city. The Lorenz asymmetry coefficient revealed the exact source of the unevenness. Spatial uncertainties were mostly caused by the tiny areas with high NDVI values, accounting for 53 cases out of 61 cases. Temporal unevenness in northwestern and northeastern parts of the lake was due to the most severe breakout occurrences, while unevenness in the lake center was mainly due to the large number of light occurrences. Finally, the advantage of Gini coefficient and Lorenz asymmetry coefficient are discussed by comparison with traditional statistical coefficients. By incorporating the two coefficients, this paper provides a quantitative and comprehensive assessment method for the spatial and temporal distribution of algal bloom.

Keywords: remote sensing, algal bloom distribution, NDVI, gini coefficient, lorenz asymmetry coefficient, chaohu lake

OPEN ACCESS

Edited by:
Zhenzhong Zeng,
Southern University of Science and Technology, China

Reviewed by:
Lian Feng,
Southern University of Science and Technology, China
Ronghua Ma,
Nanjing Institute of Geography and Limnology (CAS), China

***Correspondence:**
Ping Xia
xiaping@ahau.edu.cn

Specialty section:
This article was submitted to Environmental Informatics and Remote Sensing, a section of the journal Frontiers in Environmental Science

Received: *08 November 2021*
Accepted: *12 January 2022*
Published: *10 February 2022*

Citation:
Zhou T, Ni C, Zhang M and Xia P (2022) Assessing Spatial and Temporal Distribution of Algal Blooms Using Gini Coefficient and Lorenz Asymmetry Coefficient. Front. Environ. Sci. 10:810902. doi: 10.3389/fenvs.2022.810902

1 INTRODUCTION

Algal bloom in water areas has been a critical worldwide environment issue for the past several decades (Haag, 2007). Many studies have tracked long-term algal bloom outbreaks of various inland lakes in China (Wang et al., 2012; Yan et al., 2012; Huang et al., 2015; Zhang et al., 2015; Kang et al., 2016), U.S. (Hambright et al., 2010; Winston et al., 2014), India (Kamerosky et al., 2015), Mexico (Stumpf et al., 2003), and Canada (Hecker et al., 2012; Sorichetti et al., 2014). Since algal bloom is highly sensitive to various factors such as nutrients, temperature, wind speed, air pressure, and human controlling (Ahlgren, 1988; Vedernikov et al., 2007; Klemencic and Toman, 2010; Ribeiro et al., 2015; Wang et al., 2017), its spatial and temporal distribution is characterized as highly uneven and fluctuating, especially for inland lakes where the water flow and regeneration rate are lower. Therefore, an accurate description of spatial and temporal distribution is an important prerequisite for analyzing and controlling algal bloom outbreaks.

Satellite remote sensing has significant advantages for its large-scale and periodic observation, providing an efficient manner to observe large-scale algal blooms. Since the 1990s, the Landsat satellite has been used to monitor lake algal blooms and their dynamics (Galat and Sims, 1990; Richardson, 1996). Today there are several satellites that are widely used in algal bloom observation, i.e., the Landsat satellite (Ho et al., 2019), Sentinel satellite (Moita et al., 2016; Pirasteh et al., 2020), MODIS satellite (Lu and Tian, 2012; Zhang et al., 2015), Gaofen satellite (Hu et al., 2019), and GOCI satellite (Choi et al., 2014; Lou and Hu, 2014). Algal bloom indicators derived from satellite remote sensing bands include normalized difference vegetation index (NDVI) (Van Der Wal et al., 2010; Lin et al., 2016), FAI (Hu, 2009; Zhang et al., 2014; Page et al., 2018), Chla (Hu, 2009; Zhang et al., 2014; Page et al., 2018; Guan et al., 2020; Pompeo et al., 2021), etc. On this basis, various analyses are conducted on algal bloom distribution from spatio-temporal (Lu and Tian, 2012; Zhang et al., 2015; Page et al., 2018; Zabaleta et al., 2021) and vertical (Bosse et al., 2019) points of view. In 2019, a global spatio-temporal algal blooms analysis covering 71 large lakes from 33 countries based on Landsat five satellite images (Ho et al., 2019) revealed that algal bloom in over 2/3 of lakes had been increasing during the last 30 years. These studies show that the application of satellite remote sensing is a useful and efficient way to observe, track, and evaluate long-term and large-scale algal bloom distribution.

Although remote sensing images inversion can display the general coverage, severity, and evolution trend of algal bloom, the evaluation of distribution based on numerous images, especially for time series analysis, is subjective. Since algal bloom may grow and fade rapidly, its characteristics may be highly diverse in a couple of days (Lu and Tian, 2012; Zhang et al., 2020). Current research focuses less on the quantitative assessment method of algal bloom distributions. Using hotspots is one of the quantitative assessment methods that has been applied in spatial distribution analysis (Wei et al., 2021; Zabaleta et al., 2021). Generally, a quantitative description of temporal and spatial distribution of algal blooms is still lacking. Indices that accurately and briefly abstract the key information of algal bloom distribution features are highly needed. This requirement is more important when incorporating long-term temporal analysis in analysis.

In this paper, we focus on the assessment method of algal bloom distribution of Chaohu Lake from 2011 to 2020. Two indices, the Gini coefficient and Lorenz asymmetry coefficient, which are originally proposed to assess citizen income inequality, are applied to spatial and temporal distribution analysis of algal bloom. These two indices have been adopted to analyze river flow variability and biological species variability in previous studies (Damgaard and Weiner, 2000; Zhen-Xiang et al., 2004; Jawitz and Mitchell, 2011; Masaki et al., 2014; Zhang et al., 2020). It is considered suitable to use these two indices to measure and explain this variability. The Gini coefficient is used to measure the spatial or temporal distribution inequality (unevenness) of algal blooms, while the Lorenz asymmetry coefficient explains whether the unevenness is caused by a small number of large NDVI values or a large number of small NDVI values. To be specific, for spatial analysis, the two coefficients indicate the extent of lake-wide variability of algal bloom and which area contributes to the unevenness; for temporal analysis, the two coefficients indicate the temporal variability of algal bloom in each grid cell, and which occurrences contribute most to the unevenness.

The study site and data of Chaohu Lake are described in **section 2**. The application of the Gini coefficient and Lorenz asymmetry coefficient in assessing algal bloom spatial and temporal distribution are explained in **section 3**. **Section 4** presents the result of spatial and temporal distribution with discussion. A conclusion is given in **section 5**.

2 STUDY AREA AND DATA

2.1 Chaohu Lake

Chaohu Lake (31°25′ ~ 31°42′N, 117°17′ ~ 117°50′E), located in central-eastern China, is the fifth largest inland freshwater lake of China. The lake covers an area of 780 km^2, with a length of 55 km in longitude and a width of 21 km in latitude (Shang and Shang, 2005). A total of 90% of the Chaohu Lake is supplied by surface runoff (Yang et al., 2013), consisting of 10 major inflow-tributaries entering the lake. The location and distribution of Chaohu Lake is shown in **Figure 1**.

Algal bloom of Chaohu Lake has occurred almost every summer in the past several decades (Shang and Shang, 2005). Since the 1970s, with rapid industry and population development, nutrients and organic matter such as nitrogen and phosphorus in the lake have rapidly increased, resulting in the frequent occurrence of algal bloom (Kong et al., 2013). Satellite remote sensing data showed that algal bloom in Chaohu Lake has broken out almost every year in the past 30 years, with an average annual outbreak frequency of six times (Damgaard and Weiner, 2000). In the early 1980s, the algal blooms were mainly distributed in the northwest and northeast lake areas. During 1983–1990, the algal bloom gradually moved to the lake center. In 1990, the algal bloom

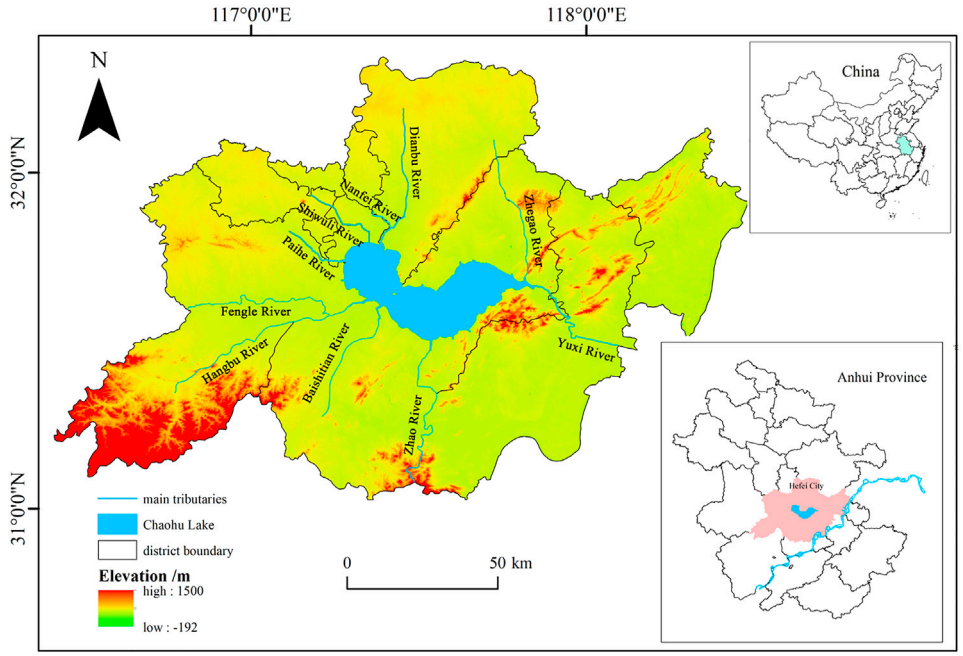


FIGURE 1 | Location and distribution of Chaohu Lake.

TABLE 1 | Parameters of Landsat satellite, GF-1 satellite, and Sentinel satellite.

Satellite	Launch year	Country	Spatial resolution[a]	Repeat period (d)
Landsat-5	1982	U.S.	30/60 m	16
Landsat-7	1999	U.S.	15/30/60 m	16
Landsat-8	2013	U.S.	15/30/100 m	16
GF-1	2013	China	16 m	4
Sentinel-2A	2015	European Space Agency	10/20/60 m	10
Sentinel-2B	2017	European Space Agency	10/20/60 m	10

[a]*Spatial resolution may vary in different bands.*

TABLE 2 | Satellite images dates and sources .

Year	Number of images	Image date and source								
2011	2	07/11[a]	09/14[a]							
2012	4	06/20[b]	07/22[b]	08/07[b]	09/24[b]					
2013	6	06/15[c]	07/09[b]	08/10[b]	08/18[c]	09/03[c]	09/19[c]			
2014	5	07/21[d]	08/05[c]	08/15[d]	08/21[c]	09/04[d]				
2015	3	08/31[d]	10/24[d]	10/27[c]						
2016	3	06/14[d]	07/25[c]	09/12[d]						
2017	2	08/25[e]	09/16[d]							
2018	13	06/06[e]	06/12[d]	07/20[d]	07/24[d]	07/29[d]	09/4[d]	09/15[d]	09/19[d]	10/2[d]
		10/03[d]	10/06[d]	10/27[d]	10/31[d]					
2019	17	05/23[d]	06/14[d]	06/26[e]	07/07[d]	07/11[e]	07/20[d]	07/21[e]	08/01[d]	08/03[c]
		09/19[d]	09/23[d]	09/29[e]	10/18[d]	10/29[e]	10/30[d]	10/31[d]	11/01[c]	
2020	6	05/17[c]	08/01[d]	08/19[e]	08/24[e]	09/03[e]	09/19[d]			
Sum	61									

[a]*: Landsat-5*
[b]*: Landsat-7*
[c]*: Landsat-8*
[d]*: GF-1;*
[e]*: Sentinel-2.*

TABLE 3 | NIR and R bands for Landsat-5, Landsat-7, Landsat-8, GF-1, and Sentinel-2 satellites.

	Landsat-5	Landsat-7	Landsat-8	GF-1	Sentinel-2
NIR	Band 4	Band 4	Band 5	Band 4	Band 8
R	Band 3	Band 3	Band 4	Band 3	Band 4

occurred throughout the lake. From 1999 to 2017, algal bloom began to shrink gradually, with most algal blooms concentrated in the northwest lake area (Li S M, 2019). The initial time of algal bloom outbreak in each year is gradually getting earlier, and the duration is gradually increasing (Damgaard and Weiner, 2000).

2.2 Data Sources

In this study, three types of satellite sensors, the Landsat 5/7/8 satellite, GF-1 satellite, and Sentinel-2 satellite are used to obtain the algal distribution of Chaohu Lake. **Table 1** lists the main parameters of the three sensors. All candidate images from May to October between the years 2011–2020 were examined, and only cloud-free images with algal bloom areas greater than 50 km^2 were considered. A total of 61 remote sensing images were obtained with significant algal bloom coverage and almost no cloud coverage. **Table 2** shows the date and source of selected images. The number of images varies widely from year to year due to the different severity of algal bloom and image quality.

All the images were checked or preprocessed to make sure atmospheric correction was applied. The aim of atmospheric correction is to eliminate the influence of atmospheric and illumination factors on the reflection of ground objects. In this study, the Landsat-5, Landsat-7, and Landsat-8 data were L2 grade and corrected by the Landsat official production system including radiometric and geometric correction (https://www.usgs.gov/faqs/does-landsat-level-1-data-processing-include-atmospheric-correction). The Sentinel-2 data (L2A level) used in this paper were generated from 1C products based on scenario classification and atmospheric correction algorithms. (https://sentinels.copernicus.eu/web/sentinel/technical-guides/sentinel-2-msi/level-2a-processing). The GF-1 data were pre-processed using FLAASH mode in ENVI 15.3 software.

2.3 NDVI Indicator

Since green algae has similar spectral characteristics with terrestrial vegetation, some indicators that reflect vegetation are widely adopted to characterize algal bloom coverage of a water area, such as FAI and NDVI (Zhu et al., 2020). In this study, due to the lack of short-wave infrared band data in the GF-1 sensor, the NDVI index was adopted as an indicator of algal bloom. The derivation of NDVI is:

$$NDVI = \frac{NIR - R}{NIR + R} \quad (1)$$

where NIR is the reflectance of the near-infrared band and R is the reflectance of the red band. The NDVI value ranges from -1 to 1. Positive NDVI denotes the existence of algal bloom, and higher NDVI indicates higher algal bloom severity. Bands of NIR and R for each satellite used in this study are listed in **Table 3**.

3 MATERIALS AND METHODS

3.1 Extraction of Lake Cells

Although the distribution of algal bloom can be roughly seen and analyzed from the NDVI distribution map of remote sensing inversion, it is still very important to divide the lake area into grid cells for quantitative assessment, because appropriate cell division can facilitate further calculation and make the spatial and temporal evaluation results closer to the real value. Grid cell division has been used in various spatial analyses for quantitative analysis. Masaki et al. (2014) extracted major river channel cells to analyze the variability of inflow regimes for different parts of rivers; Guevara-Escobar et al. (2007) used grid-divided data to evaluate rainfall distribution patterns; Raziei and Pereira (2013) and Das et al. (2014) interpolated and gridded rainfall distributions to 0.5°*0.5° and 1°*1° grid cells respectively for spatial analysis. With this consideration, the Chaohu Lake area was divided into equidistant grids for quantitative spatial and temporal analysis. First, the shape of Chaohu Lake was clipped using DEM contours. To avoid the confusion of lake edge caused by water level fluctuation, the DEM contour latitude was set slightly lower than the normal water level. The clipped lake area was then taken as the "uniform lake shape" for all images, assuring the location of grids was consistent in every image. Second, the lake area was divided by horizonal and vertical parallel lines with 0.01° distance, generating 850 0.01°*0.01° grid cells. In this way, the distribution of algal bloom was represented by uniform grid data of mean NDVI. The grid data were regarded as the regularized vector for quantitative spatial and temporal analysis. **Figure 2** shows an example of the NDVI distribution from 2014.08.15 with grid cell division.

Based on grid cell division, spatial and temporal distribution can be analyzed quantitatively and comprehensively. **Figure 3** shows an illustration of the spatial and temporal analysis based on grid cells. Spatial analysis is based on grid cells of each image, while temporal analysis is derived by the NDVI time series of each grid cell from different images.

3.2 Gini Coefficient and Lorenz Asymmetry Coefficient

In this paper, the Gini coefficient and Lorenz asymmetry coefficient are adopted to evaluate the shape of NDVI-area curve. The advantage of such coefficients is that they can describe not only the variability of NDVI over different areas but also the attribution of variability. As mentioned above, the algal bloom distribution, which is represented by NDVI, varies largely in different areas and different years in Chaohu Lake. The two indices help the interpretation of spatial and temporal inequality of algal bloom. The Gini coefficient provides the total inequality degree, while the Lorenz asymmetry coefficient interprets whether the inequality is caused by high-NDVI elements or low-NDVI elements. Considering these features of indices, we attempted to give a comprehensive description and attribution of algae distribution inequality.

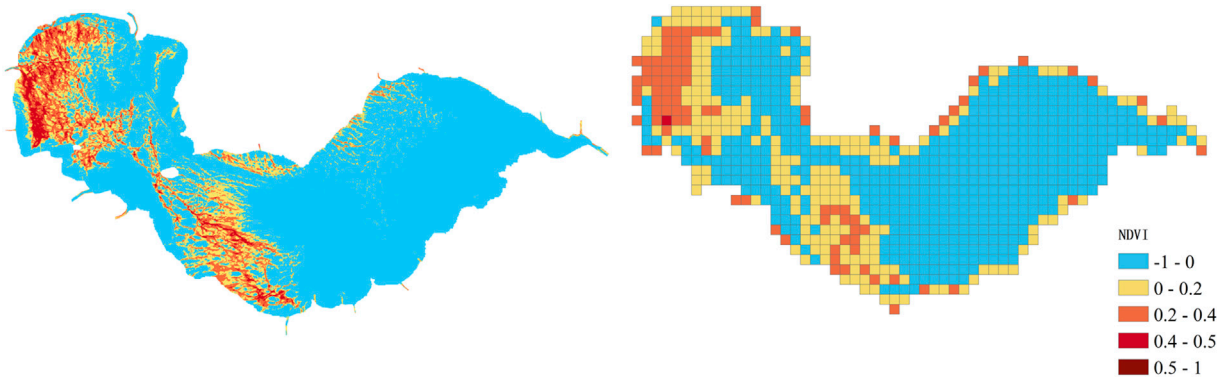


FIGURE 2 | NDVI distribution with grid cell division from 2014.08.15.

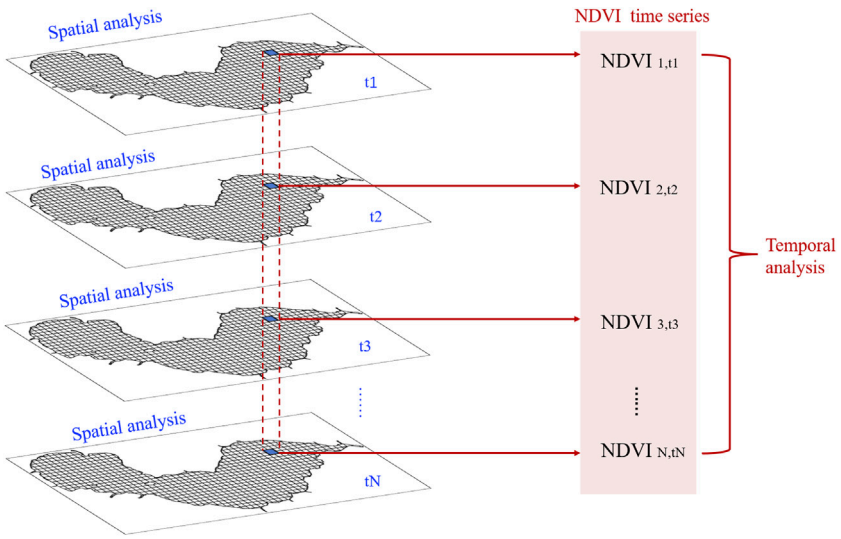


FIGURE 3 | Illustration of spatial and temporal analysis based on grid cells.

3.2.1 Gini Coefficient

The Gini coefficient was first introduced by Corrado Gini in 1912 (Gini, 1912) to quantify inequality of household incomes. In the past decades, the Gini coefficient has been extended to environment sciences, such as inequality of plant species distribution (Ma et al., 2006; Shi et al., 2012), precipitation inequality (Shi et al., 2012), and river flow temporal distribution (Masaki et al., 2014), etc. Hereafter, we apply the Gini coefficient to spatial and temporal distribution inequality of Chaohu Lake.

The Gini coefficient G is given by (Eytan, 1972; Kimura, 1994)

$$G = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\left|y_i - y_j\right|}{2n^2\bar{y}} \tag{2}$$

where n is the number of individuals, y_i is the income of individual i, and $\bar{y}$ is average income of all individuals. Note that $y_1 \leq y_2 \leq \ldots \leq y_n$. A higher Gini coefficient reflects a higher income diversity, and vice versa. When applying the Gini coefficient to assess the unevenness of algal bloom, n represents the number of total grids, y_i denotes algal bloom amount which is represented by NDVI in grid i.

The Gini coefficient is more widely interpreted using a graphical manner, known as the Lorenz curve. The Lorenz curve is obtained by aggregating the percentage of individuals (horizonal axis) and the percentage of incomes (vertical axis). In this study, the horizonal and vertical axes of the Lorenz curve are the cumulative percentage of the grid area and the cumulative percentage of algal blooms, as shown in **Figure 4**. The graphical interpretation of the Gini coefficient is the ratio of area A to the triangle (A + B) in **Figure 4**. Note that (A + B) = $\frac{1}{2}$, and the relationship between G and A is:

$$G = \frac{A}{A+B} = 2A \tag{3}$$

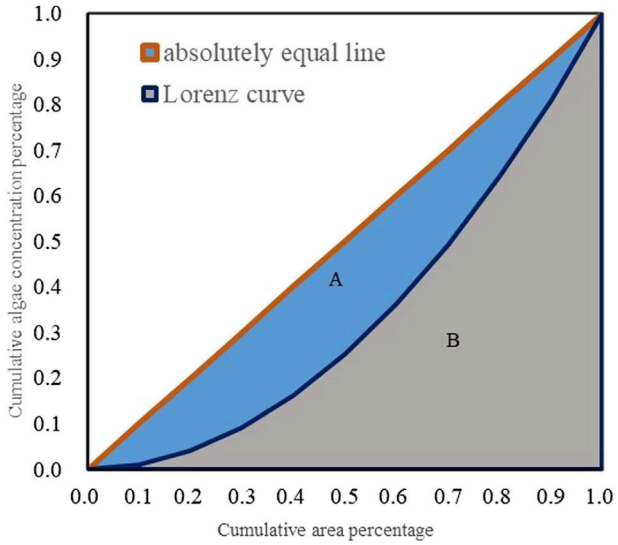


FIGURE 4 | Sketch of the Lorenz curve and Gini coefficient.

The slope of the Lorenz curve reveals the inequality degree of algal bloom spatial distribution. Imagining a perfectly even distribution of algal bloom where a 1% increase in area corresponds to a 1% increase in algal cumulation, the Lorenz curve is a line with slope 1 ($y = x$), which is called the "absolutely equal line". For unevenly distributed algal bloom area distribution, the Lorenz curve is under the absolutely equal line. The farther the Lorenz curve is from the $y = x$ line, the more uneven the algal bloom distribution is, and the higher the G value is.

3.2.2 Lorenz Asymmetry Coefficient

Although the Gini coefficient presents an efficient indicator to describe the unevenness of algae distribution, it also has some limitations. Since G = 2A, it is possible that different Lorenz curves may have the same Gini coefficient. **Figure 5** shows NDVI distributions from 2016.06.14 to 2018.09.19. Both Gini coefficients are 0.12, however their algal bloom intensity and distribution are quite diverse, also the shape of the Lorenz curves is different. This indicates that the Gini coefficient is insufficient to describe algal bloom distribution, and the Lorenz asymmetry coefficient is then adopted.

The Lorenz asymmetry coefficient, denoted as S, is given by (Damgaard and Weiner, 2000):

$$S = F + L \tag{4}$$

$$\delta = \frac{\bar{x} - x_m}{x_{m+1} - x_m} \tag{5}$$

$$F = \frac{m + \delta}{n} \tag{6}$$

$$L = \frac{\sum_{i=1}^{m} x_i + \delta x_{m+1}}{\sum_{i=1}^{n} x_i} \tag{7}$$

where m is the number of pixels with a value less than $\bar{x}$. In this paper, x denotes NDVI series for each grid. The graphical explanation to F and L is the x-coordinate and the y-coordinate of the point with slope one of the Lorenz curve. If $S = 1$, the coordinate (F, L) lies exactly on line $x + y = 1$. If $S > 1$, the coordinate (F, L) lies above line $x + y = 1$; if $S < 1$, the coordinate (F, L) lies below line $x + y = 1$. **Figure 6** shows three types of Lorenz curve with $S < 1$, $S = 1$, and $S > 1$. Triangular marks denote points with slope = 1. Note that S=F + L for the three curves in **Figure 6**: 0.59 = 0.47 + 0.12, 1 = 0.56 + 0.44, 1.11 = 0.65 + 0.46.

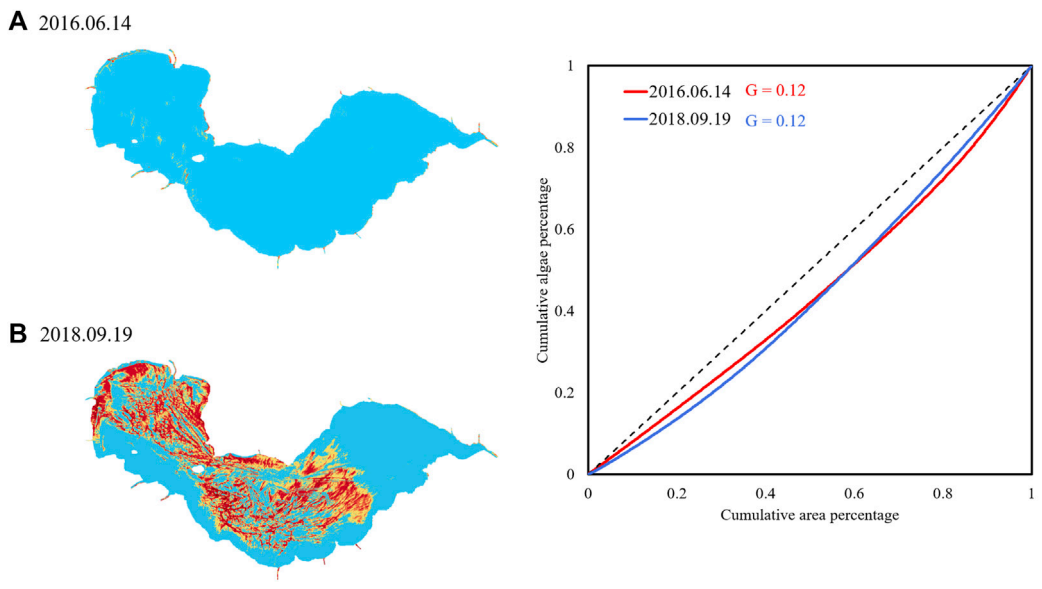


FIGURE 5 | Example of images with the same G but different algal bloom distribution and Lorenz curve.

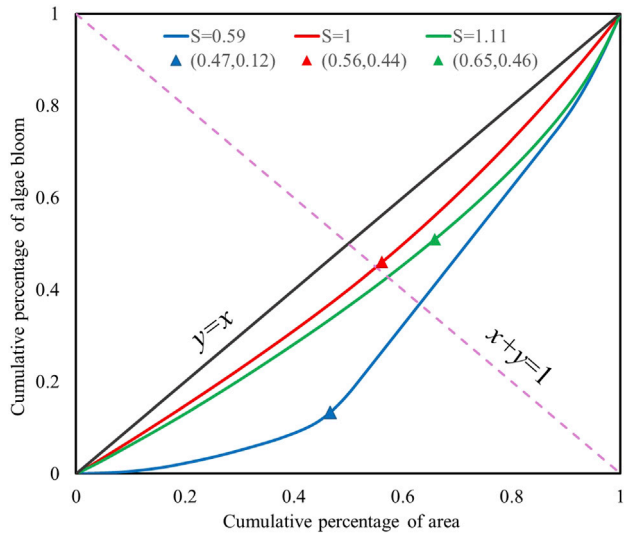


FIGURE 6 | Examples of Lorenz curves with different S values. Blue, red, and green colors indicate results for algal bloom distribution on 08/01/2020, 09/04/2014, and 10/31/2018, respectively.

The significance of the Lorenz asymmetry coefficient S is that when S > 1, the inequality of algal bloom distribution is mostly due to the small amount of high NDVI values, which is also shown in the upper-right tail of the Lorenz curve; when S < 1, the inequality of algal bloom distribution is mostly due to large amounts of low NDVI values, which is shown in the lower-left tail of the Lorenz curve; S = 1 denotes that both parts have the same contribution to algal bloom inequality. Therefore, we can infer from **Figure 6** that algal bloom unevenness on 2018.10.31 was mostly due to large NDVI elements, and unevenness on 2020.08.01 was mostly due to small NDVI elements. This provides clear notice about areas with great diversity, and, if the diversity is due to large elements (S > 1), it is an explicit warning of algal concentration.

4 RESULTS AND DISCUSSION

4.1 Spatial Distribution of NDVI in Chaohu Lake

4.1.1 General Spatial Distribution Based on Mean NDVI

The 61 remote sensing images listed in **Table 1** were processed to obtain the NDVI distribution in grid cells. **Figure 7** shows the monthly maximum NDVI from May to October. It can be clearly indicated that the algal bloom develops from May, reaching its peak in September, and then slightly decreases in October. This trend is highly accordant with the temperature trend of the Chaohu Lake area. The algal bloom first concentrates at the northwestern part of the lake (June), and then spreads to the lake center (July and August) and the whole lake (September). Besides, the lake edge and tributaries have relatively higher NDVI values than the lake center, even in May when there is almost no algal bloom throughout the lake. This indicates that algal bloom of Chaohu Lake originates from the lakeside, and is mainly imported from lake tributaries. The upper reaches of the three northwestern tributaries, Nanfei river, Paihe river, and Shiwuli river (**Figure 1**), flowing through Hefei City, which is the capital city of Anhui province with a population over 5.7 million, brings massive nutrients that trigger algal bloom at the northwestern lake area. This inference is also proved in relative studies. It is concluded that nutrient and climate conditions are two dominant issues for algal blooms of Chaohu Lake (Li et al., 2019), while Chen and Liu (2014) stated that tributaries bring 68.5% and 76.5% of nutrients (TN and TP) into Chaohu Lake; three northwestern rivers: Nanfei river, Paihe river, and Shiwuli river have the highest comprehensive pollution index among all tributaries.

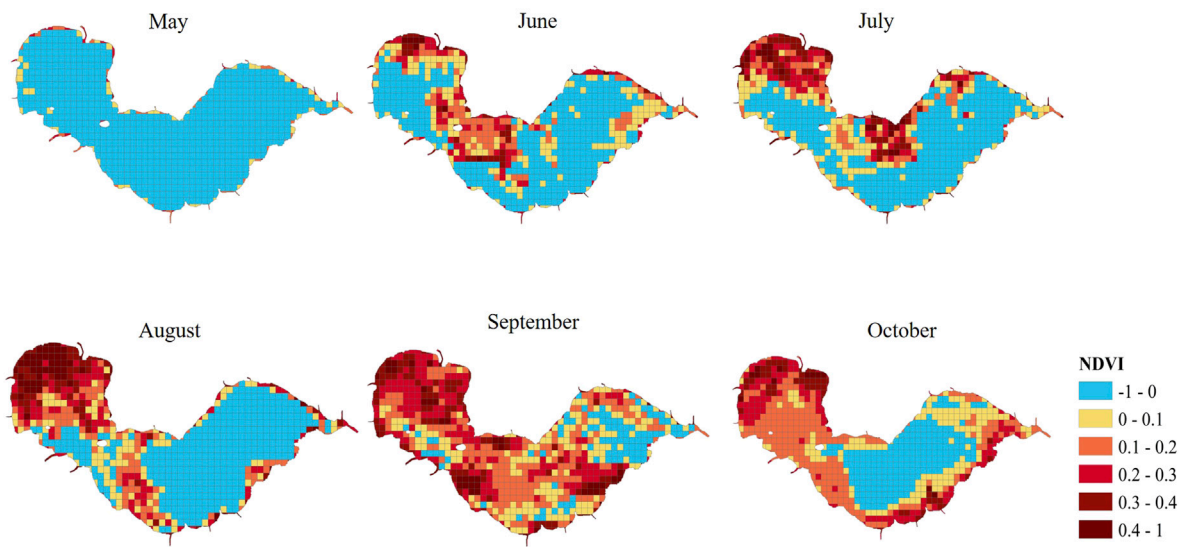


FIGURE 7 | Monthly maximum NDVI distribution of Chaohu Lake, 2011-2020.

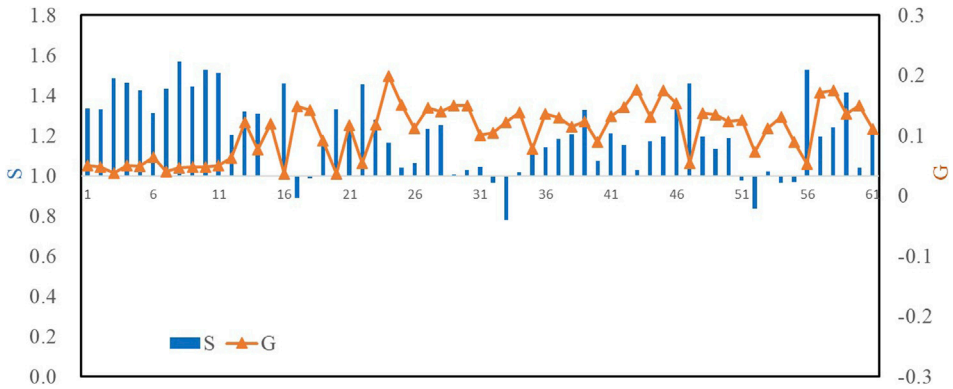


FIGURE 8 | Gini coefficients and Lorenz asymmetry coefficients from 2011 to 2020.

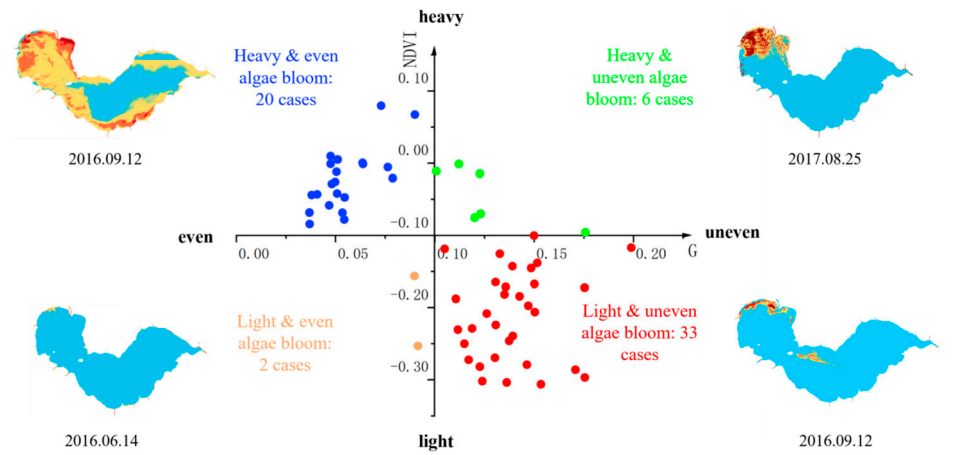


FIGURE 9 | Four types of algal bloom characterized by mean NDVI and G with example images.

4.1.2 Gini Coefficient and Lorenz Asymmetry Coefficient

1) Gini coefficient and Lorenz asymmetry coefficient results

The Gini coefficients (G) and Lorenz asymmetry coefficients (S) of 61 images are shown in **Figure 8**. G varies from 0.04 to 0.2, indicating a diverse unevenness of algal bloom. There are 53 cases with S > 1, accounting for 87% of the total 61 cases. Recalling that S implies the source of unevenness, this reveals that the algal bloom unevenness of Chaohu Lake is mostly due to the small amount of high NDVI value areas, in other word, the tiny severe algal-concentrated areas.

2) Comprehensive discussion using mean NDVI and G

By coupling mean NDVI and G, we can categorize the spatial distribution of algal bloom into four types: heavy and uneven type, heavy and even type, light and uneven type, and light and even type. **Figure 9** shows the four types with representative examples for each type. The heavy and even type and light and uneven type are in the majority, accounting for 53 cases among all 61 cases. In addition, it can be clearly found that G can be diverse in different cases even if they have the same mean NDVI, and mean NDVI can be diverse in different cases even if they have the same G. This demonstrates that univariate assessment is insufficient to describe the distribution of algal bloom. These four types of algal bloom distribution characterized by mean NDVI and G are helpful in identifying the distribution patterns of algal bloom and taking targeted measures.

3) Comprehensive discussion using mean NDVI, G, and S

In order to analyze the integrated assessment of mean NDVI, G, and S, four distribution maps are chosen from 61 images as examples: 1. minimum G (2014.08.21); 2. maximum G (2017.08.25); 3. minimum mean NDVI value (2019.05.23); and 4. maximum mean NDVI value (2019.10.29). Their NDVI distributions, Lorenz curves, mean NDVI, G, and S are shown in **Figure 10**.

The lowest and highest G (**Figure 10A**) occurred on 2014.08.21 and 2017.08.25, respectively. The NDVI map from 2014.08.21 has only sporadic high NDVI cells, emerging with an even algae distribution. In contrast, NDVI on 2017.08.25 is quite

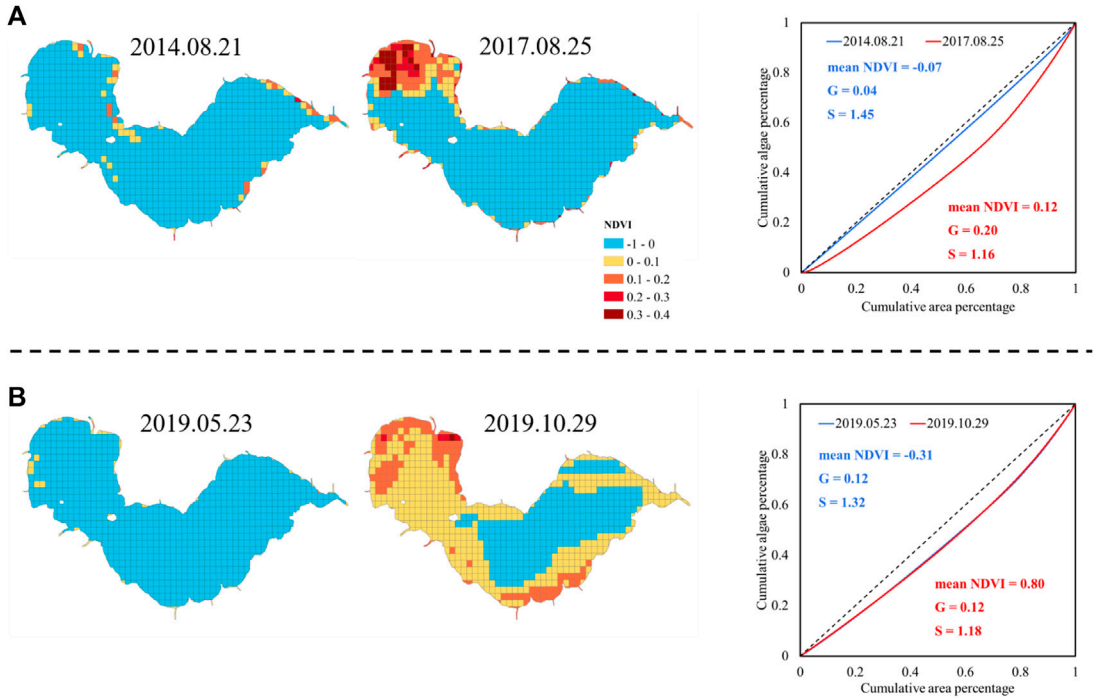


FIGURE 10 | Examples of algae distribution with **(A)** maximum and minimum G, and **(B)** maximum and minimum mean NDVI.

diverse as the northwestern part of the lake and lake side had significantly high NDVI, while NDVI in the lake center kept at a low level. This diversity is the reason for the high G value.

The minimum and maximum mean NDVI, (**Figure 10B**) which is −0.31 and 0.80, occurred on 2019.05.23 and 2019.10.29, respectively. Note that the NDVI value ranges within [−1,1], thus an average of 0.80 in NDVI indicates severe algal bloom. Algae coverage on 2019.10.29 reached 69%, while on 2019.05.23 was only 10%. However, it is interesting that their Gini coefficients are indeed the same although the NDVI difference is huge, because the Gini coefficient is related to the percentage quantiles but not the NDVI value itself.

Lorenz asymmetry coefficients (S) of the four cases are greater than 1, revealing that the unevenness of the algal bloom distribution of four cases is mainly due to the small amount of large NDVI grid cells. S value on 2014.08.21 was the greatest among the four images, revealing that the very small areas with highest NDVI in the map, are the reason for unevenness. In conclusion, mean NDVI, G, and S form a comprehensive description indicator system describing the severity and spatial distribution of algal bloom, thus providing an alternative way to quantitatively assess multiple remote sensing images.

4.2 Temporal Analysis of Algal Bloom

4.2.1 General Temporal Distribution Based on Frequency Analysis

By deriving the NDVI time series data of each grid cell, temporal analysis is carried out to reveal the change trend of algal bloom distribution for each grid cell. Here, the occurrence frequency map that consists of the frequency of each grid is derived. The frequency of grid i, F_i, is denoted by:

$$F_i = \frac{T_i}{T} \tag{8}$$

where T_i represents counts of images with $\overline{NDVI}$ >0 for grid i; T represents the number of total images, which is 61 in this study. $\overline{NDVI}$ >0 is regarded as the indication of algal bloom occurrence. The frequency map is calculated and drawn, as shown in **Figure 11**.

It is clearly indicated from **Figure 11** that the northwestern part of the lake and lake edge have significantly high frequency in algal bloom. In addition, almost all tributary estuaries (tiny branches at the lake sides) have a much higher frequency than the lake center. Similar conclusions are inferred in Zhang et al. (2015) and Liu et al. (2017) where the northwestern part of the lake has the highest frequency of algal bloom during 2000–2013, and the primary source of algal bloom is tributary and lakeshore imports.

In the northwestern part of the lake, algal bloom frequency gradually decreases from the lake side to the lake center. By reviewing NDVI distribution images, the reason is the occasional spread of algal bloom from the lake side to the lake center. Severe algal bloom may spread from the northwestern lake edge to the lake center, and the nearby northwestern lake area suffers.

4.2.2 Gini Coefficient and Lorenz Asymmetry Coefficient

1) Gini coefficient and Lorenz asymmetry coefficient results

The Gini coefficients and Lorenz asymmetry coefficients for each grid are shown in **Figure 12A** indicates that the Gini coefficient varies

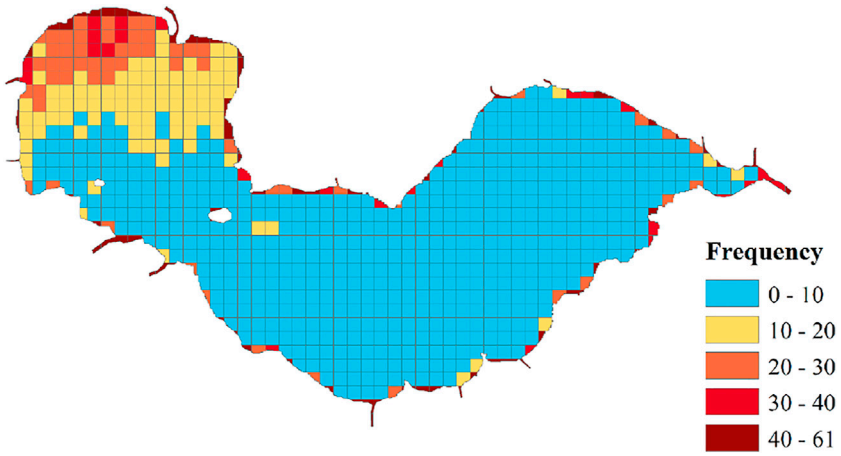


FIGURE 11 | Algal bloom frequency map for each grid cell (number of grid cells: 850).

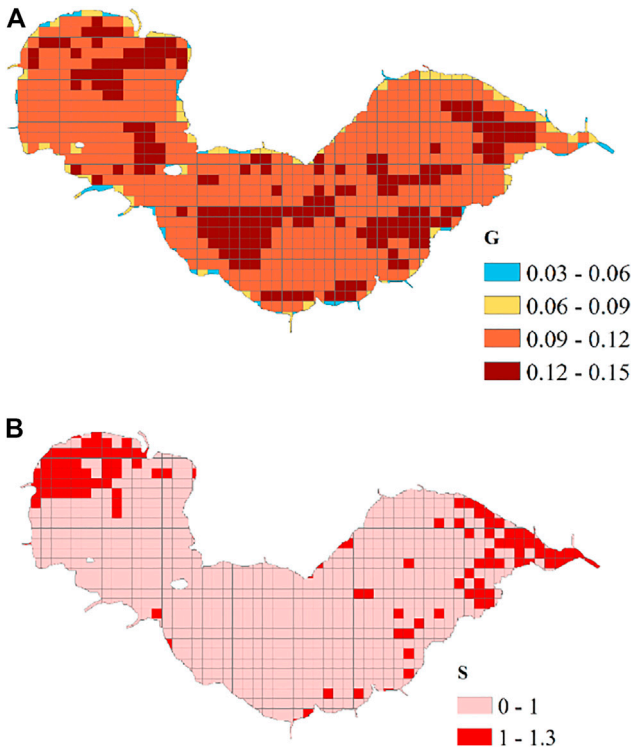


FIGURE 12 | Map of **(A)** Gini coefficient and **(B)** Lorenz asymmetry coefficient for each grid cell.

from 0.03 to 0.15 throughout the lake. In contrast with results of mean NDVI and frequency, the lake edge, where both frequency and mean NDVI are high, has the lowest G. This is because the lake edge area has a "stable and high" NDVI value during 2011–2020, and a stable NDVI level means low variance and low G. In contrast, the lake center has a relatively mixed G value, which is due to a high variance of NDVI values caused by occasional algal bloom occurrences.

Lorenz asymmetry coefficient S is divided and presented into two categories: $S > 1$ and $S < 1$, shown in **Figure 12B**. It can be clearly observed from **Figure 12B** that grid cells with $S > 1$ concentrate in the northwestern and northeastern areas, with a total of 124 grid cells. It means that the unevenness of these areas is due to the most severe occurrences among the 61 occurrences. In other words, these areas had a few extraordinarily severe algal bloom events, and these events caused the distribution unevenness in the temporal dimension.

2) Discussion on the sources of unevenness with $S > 1$.

Areas with $S > 1$ are more concerning in this study, as they imply the unevenness is due to high NDVI value occurrence which indicates severe algal bloom occurrences. Therefore, for the 124 grid cells with $S > 1$, the occurrences that contribute most to the unevenness of each grid cells are singled out. If one occurrence is responsible for multiple grid cells, the frequency is recorded by counts. We found that 39 occurrences were responsible for the unevenness of $S > 1$, as shown in **Figure 13**. Let us only take the occurrences with highest count for example, where the occurrence on 2019.10.29 involved 82 counts of all 124 grid cells. It implies that the algal bloom event on 2019.10.29 was a significant outlier that was responsible for the temporal unevenness of 2/3 grid cells. It is not surprising because the occurrence on 2019.10.29 was also mentioned in **Figure 10B** as the highest mean NDVI during 2011–2020. Also, this event was reported by the Department of Ecology and Environment of Anhui Province, and described as a "partial bloom", which is much rarer and more serious than "sporadic bloom" that often occurs (Department of Ecology and Environment of Anhui Province).

It can be inferred from the above analysis that the Gini coefficient combined with the Lorenz asymmetry coefficient is able to quantify algal bloom distribution spatially and temporally and examine the origination of unevenness by "overlapping" numerous algal bloom events. Therefore, it provides a quantitative and useful guideline for researchers and operators to rank or evaluate numerous algal bloom occurrences over time, and track influence factors for algal blooms in specific lake areas.

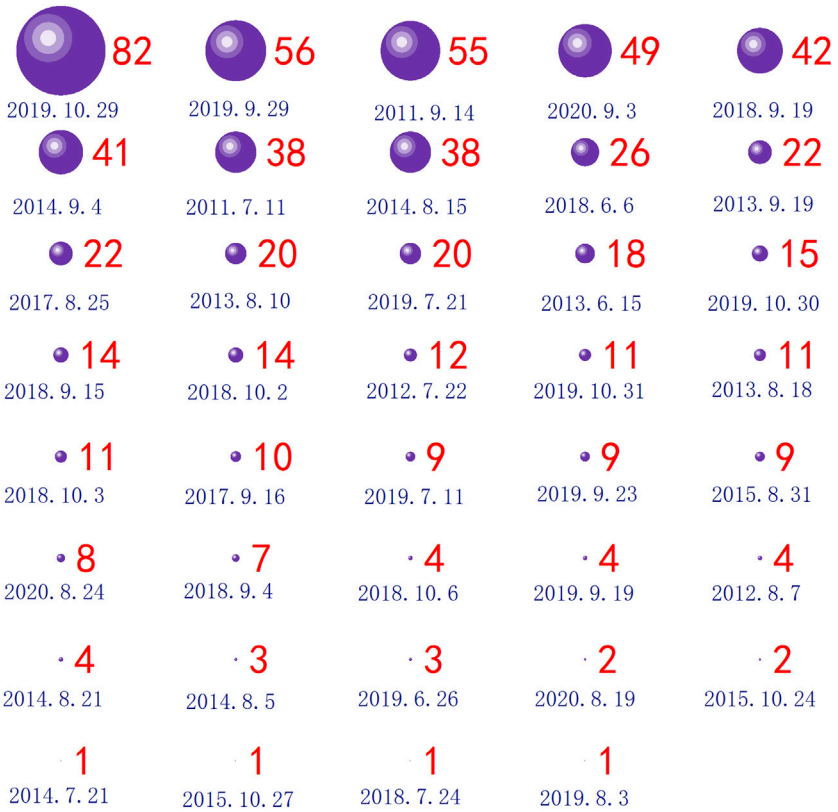


FIGURE 13 | Counts of dates that contribute to unevenness of grid cells with S > 1.

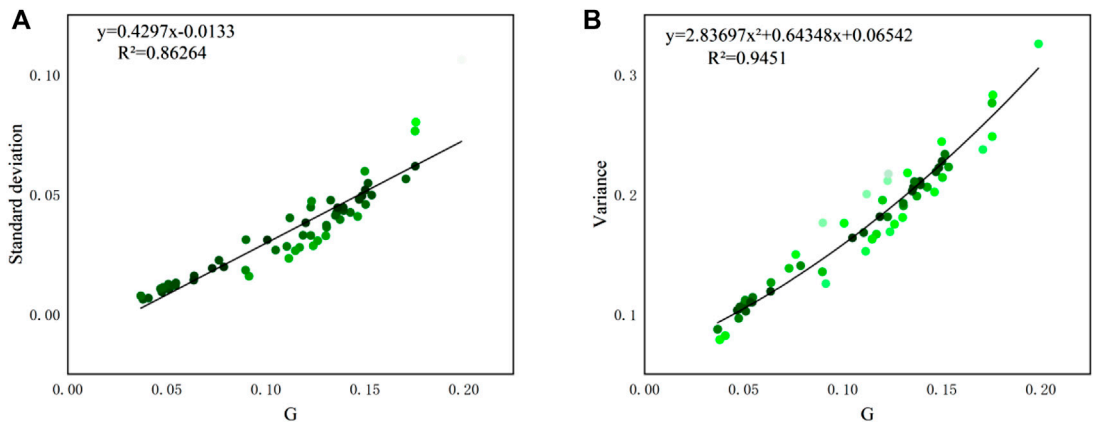


FIGURE 14 | Relationships between **(A)**. G ~ variance and **(B)**. G ~ standard deviation of spatial distribution (sample size: 61).

4.3 Comparison of the Gini Coefficient and Lorenz Curve With Other Statistical Coefficients

Since the Gini coefficient and Lorenz curve are descriptions of unevenness of distribution, there are some existing coefficients that also describe the degree of data variations, such as variance and standard deviation. As analyzed in Masaki et al. (2014) and Milanovic (1997), the Gini coefficient is proportional to the coefficient of variation and standard deviation. Also, the Lorenz curve has similarities with cumulative distribution curve (CDF) in statistics, but they have not been compared as far as we know. Hereby, the relationship and distinction between these variables are explored using the dataset in this study.

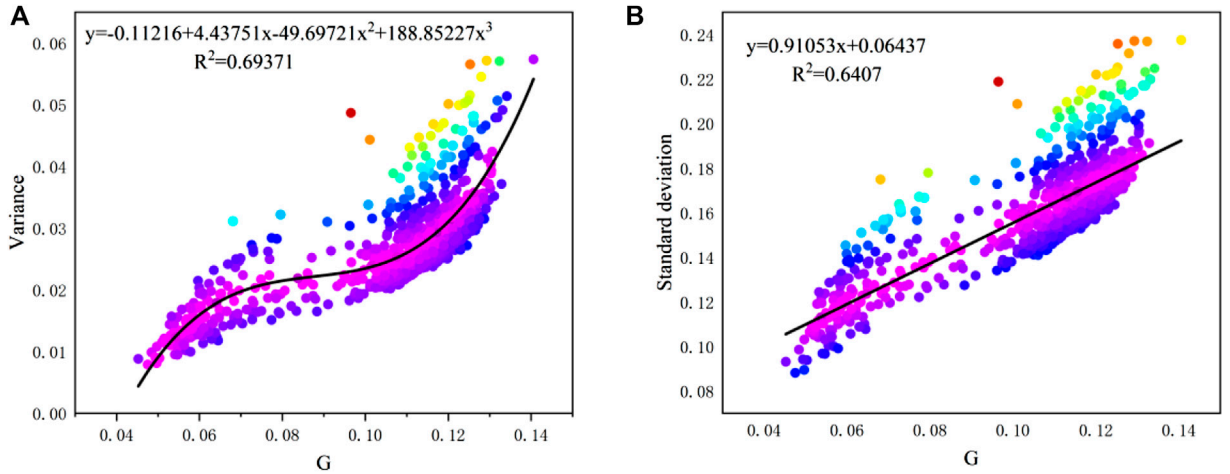


FIGURE 15 | Relationships between **(A)**. G ~ variance and **(B)**. G ~ standard deviation of temporal distribution (sample size: 850).

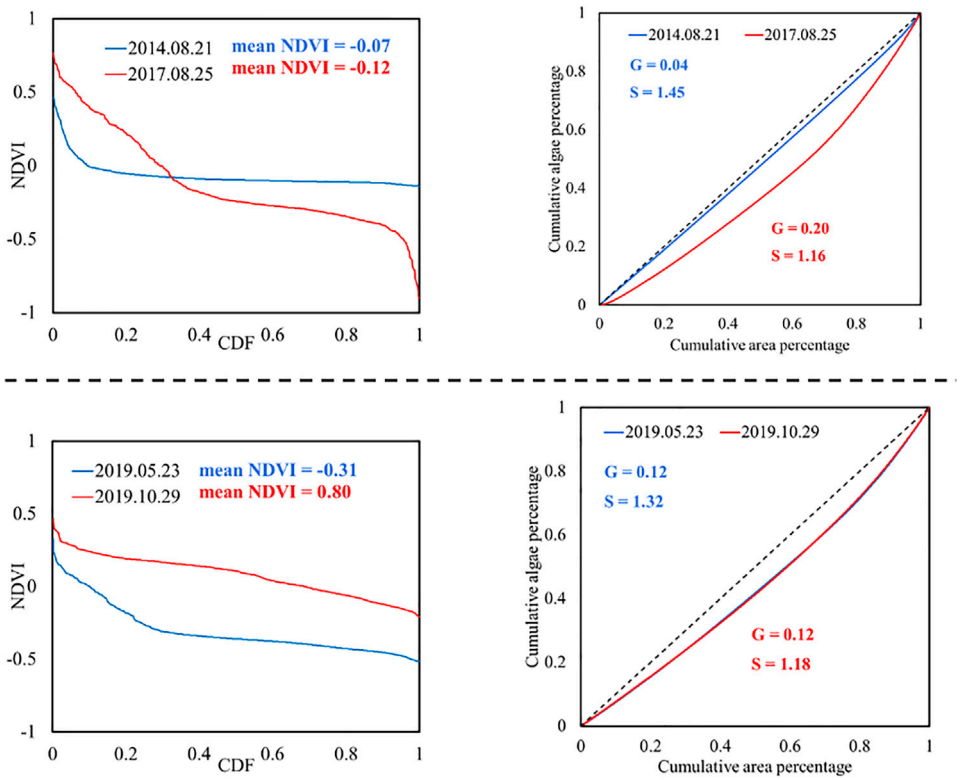


FIGURE 16 | Comparison between CDF curve and Lorenz curve.

4.3.1 Comparison Between Gini Coefficient With Other Statistical Coefficients

Fitting performance between G ~ variance and G ~ standard deviation based on spatial and temporal results are derived and shown in **Figure 14** and **Figure 15**, respectively. The results show that the Gini coefficients in this study also have fine relationships with variance and standard deviation. The regression function type is the same as the results from Masaki et al. (2014), that is, a linear relationship between G and standard deviation, and polynomial relationship between G and variance. Nevertheless, the coefficient G has an outstanding feature over the other two coefficients: the G value is normalized to [0,1] regardless of the

value of samples data, which allows it to be used as a universal indicator for all cases.

4.3.2 Comparison Between Lorenz Curve With CDF Curve

The CDF curve, which is derived by aggregating the NDVI value in ascending order and its corresponding non-exceedance probability, can also ascertain the variability of samples in the form of curves. Here, the four NDVI distributions in **Figure 10** are taken as a study example to compare the performances of CDF and Lorenz curves. **Figure 16** shows the results of CDF and Lorenz curves for each distribution.

Figure 16 shows that both the CDF curve and Lorenz curve can reveal the variability features of spatial distributions. However, the CDF curve has similar limitations with variance and standard deviation in that it is not a universal coefficient, for its vertical axis (NDVI) varies with the NDVI values. The Lorenz curve, however, has an axis of cumulative percentage that is restricted within [0, 1]. Besides, the features of the Lorenz curve can be interpreted by the Gini coefficient and Lorenz asymmetry coefficient. By these two coefficients we can easily judge various distributions without comparing the curves. However, it is not convenient to compare various CDF curves as there is no such scalar coefficient to describe curve features. Although the CDF curves in **Figure 16** are very diverse and it is easy to tell the difference, it could be confusing in comparing various sample sets with similar distributions. In conclusion, the Lorenz curve outperforms the CDF curve in assessment analysis for its regularization and comparability.

5 CONCLUSION

This paper examined the characteristics of algal bloom distribution between 2011 and 2020 using mean NDVI, Gini coefficient, and Lorenz asymmetry coefficient of Chaohu lake, China. By dividing 61 remote sensing images into equidistant grid cells, statistical analysis can be carried out based on grid cell data to explore spatial and temporal distribution and trend in a quantitative way. Results suggest that algal bloom is severe at the lake edge and northwestern part of Chaohu Lake owing to tributary and lakeside nutrient imports. Lorenz asymmetry coefficient is applied to detect the source of unevenness, and the primary source of spatial distribution unevenness is the small area with very high NDVI values. Temporal analysis shows that the northwestern part and lake edge has very high algal bloom frequency but low Gini coefficient, indicating a stable and severe algal bloom situation. Lorenz asymmetry coefficient reveals that unevenness in the 124 grid cells concentrated in the northwestern and eastern parts of the lake is due to the most severe algal bloom occurrences.

Analysis in this paper indicates that mean NDVI, Gini coefficient, and Lorenz asymmetry coefficient can comprehensively and quantitively describe the distribution characteristics of lake algal bloom, while any single coefficient is one-sided and insufficient to accurately depict the distribution information. The algal bloom distribution may be different even if they have the same mean NDVI or Gini coefficient. The compound assessment method could allow researchers to identify algal bloom distribution patterns as well as sources to the unevenness. Possible extensions of this work will include the analysis of connections between algal booms and meteorological factors for the selected extreme occurrences, and the application of the methodology to assessment in other spatial distributions.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

TZ wrote the paper and analyzed the results. CN and MZ calculated the Gini coefficients and Lorenz asymmetry coefficients, and concluded literature review for this paper. PX designed the research idea and layout of paper.

FUNDING

This research was funded by the National Science and Technology Major Project of China (grant no. 2017ZX07603-002) and the Natural Science Fund of Anhui Province (grant nos. 2008085ME158 and 2008085ME159).

REFERENCES

Ahlgren, G. (1988). Phosphorus as Growth-Regulating Factor Relative to Other Environmental Factors in Cultured Algae. *Hydrobiologia* 170, 191–210. doi:10.1007/bf00024905

Bosse, K. R., Sayers, M. J., Shuchman, R. A., Fahnenstiel, G. L., Ruberg, S. A., Fanslow, D. L., et al. (2019). Spatial-temporal Variability of *In Situ* Cyanobacteria Vertical Structure in Western Lake Erie: Implications for Remote Sensing Observations. *J. Great Lakes Res.* 45, 480–489. doi:10.1016/j.jglr.2019.02.003

Chen, Y.-Y., and Liu, Q.-Q. (2014). On the Horizontal Distribution of Algal-Bloom in Chaohu Lake and its Formation Process. *Acta Mech. Sin.* 30, 656–666. doi:10.1007/s10409-014-0078-x

Choi, J.-K., Min, J.-E., Noh, J. H., Han, T.-H., Yoon, S., Park, Y. J., et al. (2014). Harmful Algal Bloom (HAB) in the East Sea Identified by the Geostationary Ocean Color Imager (GOCI). *Harmful Algae* 39, 295–302. doi:10.1016/j.hal.2014.08.010

Damgaard, C., and Weiner, J. (2000). Describing Inequality in Plant Size or Fecundity. *Ecology* 81, 1139–1142. doi:10.1890/0012-9658(2000)081[1139:diipso]2.0.co;2

Das, P. K., Chakraborty, A., and Seshasai, M. V. R. (2014). Spatial Analysis of Temporal Trend of Rainfall and Rainy Days during the Indian Summer Monsoon Season Using Daily Gridded (0.5° × 0.5°) Rainfall Data for the Period of 1971-2005. *Met. Apps* 21, 481–493. doi:10.1002/met.1361

Department of Ecology and Environment of Anhui Province (2019). *Monitoring results of emergency prevention and control of cyanobacteria in Chaohu Lake from October 28 to November 3, 2019*. Available at: https://sthjt.ah.gov.cn/public/21691/110831651.html (Accessed November 13, 2019) (In Chinese).

Eytan, S. (1972). Relation between a Social Welfare Function and the Gini index of Income Inequality. *J. Econ. Theor.* 4, 98–100. doi:10.1016/0022-0531(72)90167-6

Galat, D. L., Verdin, J. P., and Sims, L. L. (1990). Large-scale Patterns of Nodularia Spumigena Blooms in Pyramid Lake, Nevada, Determined from Landsat Imagery: 1972?1986. *Hydrobiologia* 197, 147–164. doi:10.1007/bf00026947

Gini, C. (1912). Variabilità e mutabilità. *Contributo allo Studio delle Distribuzioni e delle Relazioni Statistiche.* Bologna: C. Cuppini

Guan, Q., Feng, L., Hou, X. J., Schurgers, G., Zheng, Y., and Tang, J. (2020). Eutrophication Changes in Fifty Large Lakes on the Yangtze Plain of China Derived from MERIS and OLCI Observations. *Remote Sens. Environ.* 246, 111890. doi:10.1016/j.rse.2020.111890

Guevara-Escobar, A., González-Sosa, E., Véliz-Chávez, C., Ventura-Ramos, E., and Ramos-Salinas, M. (2007). Rainfall Interception and Distribution Patterns of Gross Precipitation Around an Isolated Ficus Benjamina Tree in an Urban Area. *J. Hydrol.* 333, 532–541. doi:10.1016/j.jhydrol.2006.09.017

Haag, A. L. (2007). Algal Bloom Again. *Nature* 447, 520–521. doi:10.1038/447520a

Hambright, K. D., Zamor, R. M., Easton, J. D., Glenn, K. L., Remmel, E. J., and Easton, A. C. (2010). Temporal and Spatial Variability of an Invasive Toxigenic Protist in a North American Subtropical Reservoir. *Harmful Algae* 9, 568–577. doi:10.1016/j.hal.2010.04.006

Hecker, M., Khim, J. S., Giesy, J. P., Li, S.-Q., and Ryu, J.-H. (2012). Seasonal Dynamics of Nutrient Loading and Chlorophyll A in a Northern Prairies Reservoir, Saskatchewan, Canada. *J. Water Resource Prot.* 04, 180–202. doi:10.4236/jwarp.2012.44021

Ho, J. C., Michalak, A. M., and Pahlevan, N. (2019). Widespread Global Increase in Intense lake Phytoplankton Blooms since the 1980s. *Nature* 574, 667–670. doi:10.1038/s41586-019-1648-7

Hu, L., Zeng, K., Hu, C., and He, M.-X. (2019). On the Remote Estimation of Ulva Prolifera Areal Coverage and Biomass. *Remote Sensing Environ.* 223, 194–207. doi:10.1016/j.rse.2019.01.014

Hu, C. (2009). A Novel Ocean Color index to Detect Floating Algae in the Global Oceans. *Remote Sensing Environ.* 113, 2118–2129. doi:10.1016/j.rse.2009.05.012

Huang, C., Shi, K., Yang, H., Li, Y., Zhu, A.-x., Sun, D., et al. (2015). Satellite Observation of Hourly Dynamic Characteristics of Algae with Geostationary Ocean Color Imager (GOCI) Data in Lake Taihu. *Remote Sensing Environ.* 159, 278–287. doi:10.1016/j.rse.2014.12.016

Jawitz, J. W., and Mitchell, J. (2011). Temporal Inequality in Catchment Discharge and Solute export. *Water Resour. Res.* 47, W00J14. doi:10.1029/2010wr010197

Kamerosky, A., Cho, H., and Morris, L. (2015). Monitoring of the 2011 Super Algal Bloom in Indian River Lagoon, FL, USA, Using MERIS. *Remote Sensing* 7, 1441–1460. doi:10.3390/rs70201441

Kang, L., He, Q.-S., He, W., Kong, X.-Z., Liu, W.-X., Wu, W.-J., et al. (2016). Current Status and Historical Variations of DDT-Related Contaminants in the Sediments of Lake Chaohu in China and Their Influencing Factors. *Environ. Pollut.* 219, 883–896. doi:10.1016/j.envpol.2016.08.072

Kimura, K. (1994). A Micro-macro Linkage in the Measurement of Inequality: Another Look at the Gini Coefficient. *Qual. Quant.* 28, 83–97. doi:10.1007/bf01098727

Klemencic, A. K., and Toman, M. J. (2010). Influence of Environmental Variables on Benthic Algal Associations from Selected Extreme Environments in Slovenia in Relation to the Species Identification. *Period. Biol.* 112, 179–191. doi:10.4028/www.scientific.net/AMM.481.235

Kong, X.-Z., Jørgensen, S. E., He, W., Qin, N., and Xu, F.-L. (2013). Predicting the Restoration Effects by a Structural Dynamic Approach in Lake Chaohu, China. *Ecol. Model.* 266, 73–85. doi:10.1016/j.ecolmodel.2013.07.001

Li, S. M., Song, K. S., Liang, C., and Gao, J. (2019). Analysis on Spatial and Temporal Character of Algae Bloom in lake Chaohu and its Dring Factors Based on Landsat Imagery. *Resouces Environ. Yangtze Basin* 28, 1205–1213.

Lin, Y., Ye, Z., Zhang, Y., and Yu, J. (2016). Spectral Feature Analysis for Quantitative Estimation of Cyanobacteria Chlorophyll-A. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* XLI-B7, 91–98. doi:10.5194/isprsarchives-xli-b7-91-2016

Liu, C., Zhang, L., Fan, C., Xu, F., Chen, K., and Gu, X. (2017). Temporal Occurrence and Sources of Persistent Organic Pollutants in Suspended Particulate Matter from the Most Heavily Polluted River Mouth of Lake Chaohu, China. *Chemosphere* 174, 39–45. doi:10.1016/j.chemosphere.2017.01.082

Lou, X., and Hu, C. (2014). Diurnal Changes of a Harmful Algal Bloom in the East China Sea: Observations from GOCI. *Remote Sensing Environ.* 140, 562–572. doi:10.1016/j.rse.2013.09.031

Lu, C., and Tian, Q. (2012). Extracting Temporal and Spatial Distributions Information about Algal Glooms Based on Multitemporal Modis. *Anim. Genet.* 46, 220–223. doi:10.5194/isprsarchives-xxxix-b7-131-2012

Ma, Z., Shi, J., Wang, G., and He, Z. (2006). Temporal Changes in the Inequality of Early Growth of Cunninghamia Lanceolata (Lamb.) hook.: A Novel Application of the Gini Coefficient and Lorenz Asymmetry. *Genetica* 126, 343–351. doi:10.1007/s10709-005-1358-y

Masaki, Y., Hanasaki, N., Takahashi, K., and Hijioka, Y. (2014). Global-scale Analysis on Future Changes in Flow Regimes Using Gini and Lorenz Asymmetry Coefficients. *Water Resour. Res.* 50, 4054–4078. doi:10.1002/2013wr014266

Milanovic, B. (1997). A Simple Way to Calculate the Gini Coefficient, and Some Implications. *Econ. Lett.* 56, 45–49. doi:10.1016/s0165-1765(97)00101-8

Moita, M. T., Pazos, Y., Rocha, C., Nolasco, R., and Oliveira, P. B. (2016). Toward Predicting Dinophysis Blooms off NW Iberia: A Decade of Events. *Harmful Algae* 53, 17–32. doi:10.1016/j.hal.2015.12.002

Page, B. P., Kumar, A., and Mishra, D. R. (2018). A Novel Cross-Satellite Based Assessment of the Spatio-Temporal Development of a Cyanobacterial Harmful Algal Bloom. *Int. J. Appl. Earth Obs. Geoinf.* 66, 69–81. doi:10.1016/j.jag.2017.11.003

Pirasteh, S., Mollaee, S., Fatholahi, S. N., and Li, J. (2020). Estimation of Phytoplankton Chlorophyll-A Concentrations in the Western Basin of Lake Erie Using Sentinel-2 and Sentinel-3 Data. *Can. J. Remote Sensing* 46, 585–602. doi:10.1080/07038992.2020.1823825

Pompêo, M., Moschini-Carlos, V., Bitencourt, M. D., Sòria-Perpinyà, X., Vicente, E., and Delegido, J. (2021). Water Quality Assessment Using Sentinel-2 Imagery with Estimates of Chlorophyll a, Secchi Disk Depth, and Cyanobacteria Cell Number: the Cantareira System Reservoirs (São Paulo, Brazil). *Environ. Sci. Pollut. Res.* 28, 34990–35011. doi:10.1007/s11356-021-12975-x

Raziei, T., and Pereira, L. S. (2013). Spatial Variability Analysis of Reference Evapotranspiration in Iran Utilizing fine Resolution Gridded Datasets. *Agric. Water Manag.* 126, 104–118. doi:10.1016/j.agwat.2013.05.003

Ribeiro, F., Gallego-Urrea, J. A., Goodhead, R. M., Van Gestel, C. A. M., Moger, J., Soares, A. M. V. M., et al. (2015). Uptake and Elimination Kinetics of Silver Nanoparticles and Silver Nitrate byRaphidocelis Subcapitata: The Influence of Silver Behaviour in Solution. *Nanotoxicology* 9, 686–695. doi:10.3109/17435390.2014.963724

Richardson, L. L. (1996). Remote Sensing of Algal Bloom Dynamics: New Research Fuses Remote Sensing of Aquatic Ecosystems with Algal Accessory Pigment Analysis. *Bioscience* 46, 492–501.

Shang, G.-p., and Shang, J.-c. (2005). Causes and Control Countermeasures of Eutrophication in Chaohu Lake, China. *Chin. Geograph. Sc.* 15, 348–354. doi:10.1007/s11769-005-0024-8

Shi, W. L., Yang, Q. K., Li, X. F., Chen, A., and He, X. J. (2012). Study on Temporal Inequality of Precipitation in the Loess Plateau Based on Lorenz Curve. *Agric. Res. Arid Areas* 30, 172–177. doi:10.3969/j.issn.1000-7601.2012.04.031

Sorichetti, R. J., Mclaughlin, J. T., Creed, I. F., and Trick, C. G. (2014). Suitability of a Cytotoxicity Assay for Detection of Potentially Harmful Compounds Produced by Freshwater Bloom-Forming Algae. *Harmful Algae* 31, 177–187. doi:10.1016/j.hal.2013.11.001

Stumpf, R. P., Culver, M. E., Tester, P. A., Tomlinson, M., Kirkpatrick, G. J., Pederson, B. A., et al. (2003). Monitoring Karenia Brevis Blooms in the Gulf of Mexico Using Satellite Ocean Color Imagery and Other Data. *Harmful Algae* 2, 147–160. doi:10.1016/s1568-9883(02)00083-5

Van Der Wal, D., Wielemaker-Van Den Dool, A., and Herman, P. M. J. (2010). Spatial Synchrony in Intertidal Benthic Algal Biomass in Temperate Coastal and Estuarine Ecosystems. *Ecosystems* 13, 338–351. doi:10.1007/s10021-010-9322-9

Vedernikov, V. I., Bondur, V. G., Vinogradov, M. E., Landry, M. R., and Tsidilina, M. N. (2007). Anthropogenic Influence on the Planktonic Community in the basin of Mamala Bay (Oahu Island, Hawaii) Based on Field and Satellite Data. *Oceanology* 47, 221–237. doi:10.1134/s0001437007020099

Wang, S., Zhang, M., Li, B., Xing, D., Wang, X., Wei, C., et al. (2012). Comparison of Mercury Speciation and Distribution in the Water Column and Sediments between the Algal Type Zone and the Macrophytic Type Zone in a Hypereutrophic lake (Dianchi Lake) in Southwestern China. *Sci. Total Environ.* 417-418, 204–213. doi:10.1016/j.scitotenv.2011.12.036

Wang, L., Wang, X., Jin, X., Xu, J., Zhang, H., Yu, J., et al. (2017). Analysis of Algae Growth Mechanism and Water Bloom Prediction under the Effect of Multi-Affecting Factor. *Saudi J. Biol. Sci.* 24, 556–562. doi:10.1016/j.sjbs.2017.01.026

Wei, X. D., Wang, N., Luo, P. P., Yang, J., Zhang, J., and Lin, K. L. (2021). Spatiotemporal Assessment of Land Marketization and its Driving Forces for Sustainable Urban-Rural Development in Shaanxi Province in China. *Sustainability* 13, 7755. doi:10.3390/su13147755

Winston, B., Hausmann, S., Scott, J. T., and Morgan, R. (2014). The Influence of Rainfall on Taste and Odor Production in a South-central USA Reservoir. *Freshw. Sci.* 33, 755–764. doi:10.1086/677176

Yan, Q., Li, Y., Huang, B., Wang, A., Zou, H., Miao, H., et al. (2012). Proteomic Profiling of the Acid Tolerance Response (ATR) during the Enhanced Biomethanation Process from Taihu Blue Algae with Butyrate Stress on Anaerobic Sludge. *J. Hazard. Mater.* 235-236, 286–290. doi:10.1016/j.jhazmat.2012.07.062

Yang, L., Lei, K., Meng, W., Fu, G., and Yan, W. (2013). Temporal and Spatial Changes in Nutrients and Chlorophyll-α in a Shallow lake, Lake Chaohu, China: An 11-year Investigation. *J. Environ. Sci.* 25, 1117–1123. doi:10.1016/s1001-0742(12)60171-5

Zabaleta, B., Achkar, M., and Aubriot, L. (2021). Hotspot Analysis of Spatial Distribution of Algae Blooms in Small and Medium Water Bodies. *Environ. Monit. Assess.* 193, 221. doi:10.1007/s10661-021-08944-z

Zhang, Y., Ma, R., Duan, H., Loiselle, S. A., Xu, J., and Ma, M. (2014). A Novel Algorithm to Estimate Algal Bloom Coverage to Subpixel Resolution in Lake Taihu. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sensing* 7, 3060–3068. doi:10.1109/jstars.2014.2327076

Zhang, Y., Ma, R., Zhang, M., Duan, H., Loiselle, S., and Xu, J. (2015). Fourteen-Year Record (2000-2013) of the Spatial and Temporal Dynamics of Floating Algae Blooms in Lake Chaohu, Observed from Time Series of MODIS Images. *Remote Sensing* 7, 10523–10542. doi:10.3390/rs70810523

Zhang, Y., Luo, P., Zhao, S., Kang, S., Wang, P., Zhou, M., et al. (2020). Control and Remediation Methods for Eutrophic Lakes in the Past 30 Years. *Water Sci. Technol.* 81, 1099–1113. doi:10.2166/wst.2020.218

Zhen-Xiang, H., Pei, Q., Cheng-Jiang, R., Min, X., and Mopper, S. (2004). Lorenz Curve and its Application in Plant Ecology. *J. Nanjing For. Univ.* 28, 37–41. doi:10.17521/cjpe.2004.0086

Zhu, Y. H., Luo, P. P., Zhang, S., and Sun, B. (2020). Spatiotemporal Analysis of Hydrological Variations and Their Impacts on Vegetation in Semiarid Areas from Multiple Satellite Data. *Remote Sensing* 12, 4177. doi:10.3390/rs12244177

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Environmental Science

ORIGINAL RESEARCH
published: 25 February 2022
doi: 10.3389/fenvs.2022.819693

Evaluation of Water Resources Utilization Efficiency in Guangdong Province Based on the DEA–Malmquist Model

Lifen Cheng[1], *Song Song*[1,2]* *and Yufeng Xie*[1]

[1]School of Geography and Remote Sensing, Guangzhou University, Guangzhou, China, [2]Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), Guangzhou, China

In this study, we examined the temporal and spatial trends of water utilization efficiency in the highly developed Guangdong Province based on a data envelopment analysis (DEA) model and Malmquist index, trying to evaluate and predict the water utilization status in prefecture-level cities from the input–output perspective. The results showed that the water utilization efficiency in Guangdong Province in 2012–2018 was on a fluctuating upward trend, although the utilization efficiency had not reached the DEA efficiency; spatially, the water utilization efficiency in Guangdong Province was relatively high in the Pearl River Delta and the surrounding cities, but low in the Western Guangdong region; in the perspective of efficiency decomposition factors, technological change was the most pervasive driving force of the water utilization efficiency. Pearl River Delta and Northern Guangdong made progress in productivity due to the advantages of scale expansion benefit. Eastern Guangdong achieved improvement in all-over parameters and made the largest growth in water utilization efficiency, while Western Guangdong basically profited from organizational management during the study period; the prediction analysis revealed a continuous increase in the water utilization efficiency in Guangdong Province during 2019–2025 with an average annual growth rate of about 0.6%, benefiting mainly from the technological innovation and secondly from management renewal, while scale expansion benefit tended to decline. Furthermore, spatial heterogeneity exhibited a decreasing tendency in the future owing to the rapid water utilization efficiency increase in the Western and Eastern Guangdong. It is suggested that the technological innovation and the integrated management capability in water use should be paid more attention to enhance the water utilization efficiency in Guangdong Province. Achieving a balance between economic growth and water resource protection, and promoting the sustainable development of the nature-economy-society compound system is of the utmost priority in Guangdong.

Keywords: water utilization efficiency, evaluation, prediction, DEA model, malmquist index model

OPEN ACCESS

Edited by:
Chuanfu Zang,
South China Normal University, China

Reviewed by:
Mario Martín Gamboa,
Rey Juan Carlos University, Spain
Ehsan Elahi,
Shandong University of Technology, China

***Correspondence:**
Song Song
geossong@gzhu.edu.cn

Specialty section:
This article was submitted to Environmental Economics and Management, a section of the journal Frontiers in Environmental Science

Received: *22 November 2021*
Accepted: *10 January 2022*
Published: *25 February 2022*

Citation:
Cheng L, Song S and Xie Y (2022) Evaluation of Water Resources Utilization Efficiency in Guangdong Province Based on the DEA–Malmquist Model. Front. Environ. Sci. 10:819693. doi: 10.3389/fenvs.2022.819693

INTRODUCTION

Water is the fundamental resource for the living organism, and the indispensable ingredient of animate and non-animate systems on earth. Water resources serve as the basic element to the national social-economic production, and at the same time support the healthy evolution and beneficent circulation of ecological processes. Unreasonable water use will inevitably bring restrictions on ecosystem health, socioeconomic development, and even human survival. Water shortage has become a bottleneck problem restricting economic development and influencing the population's livelihood in China and the world (Chen et al., 2020). It was estimated that more than two thirds of cities in China are in short supply of water resources, while half of them are confronting serious water shortages (Wan and Zhang, 2012), while globally around 4 billion people are bedeviled by severe water shortages (Khaskheli et al., 2021; Salehi, 2022). In the past decades, China has been suffering a series water related stress, including a sharp water demand surge, rapid water quality deterioration, and continuous water environment degradation due to long-term and high-intensity exploitation of water resources. Considering the integrated impact of climate change, population growth, and industrialization, the conflict between water demand and supply will be more acute, especially in the highly developed regions.

The United Nations' 2030 Agenda for Sustainable Development clearly states the direction for the sustainable development of water resources, by emphasizing that "every country need to ensure the sustainable supply of fresh water, greatly reduce the number of people suffering from water stress, and improve water-use efficiency by 2030, " to alleviate the ever-growing pressure in water resources and water environments (Adeel, 2017). Accurate evaluation and gradual improvement of the water utilization efficiency (WUE) are the basic steps to facilitate the sustainable development of water resources. The WUE is concomitant with labor and capital (Ding et al., 2019), and is defined as the economic value of products manufactured per unit of water resources consumed (Liu et al., 2020) WUE is closely associated with a country's economic strength and scientific-technical advancement. The United Nations has listed China as one of the 13 countries with severe water shortage, due to the low per capita water availability (less than one-quarter of the world's average level) and the inferior WUE compared with those of other countries. In the United States, however, the development, utilization, and management of water resources have reached the world's leading level since the 1930s (Fan et al., 2008). How to improve the WUE has become one of the primary concern of researchers, policy makers, and business leaders to relieve China's water resource crisis.

Researchers in China and worldwide have primarily evaluated WUE on a variety of spatial and temporal scales. Existing research mainly focuses on the comparation of WUE in different industrial sectors and the suitability of various evaluation methods (García-Valiñas and Muñiz, 2007; Sun and Yang, 2019; Qi and Song, 2020; Shi et al., 2021). WUE showed significant variation among regions and cities from different perspectives, and wide fluctuation over time can also be witnessed due to the productivity and economic structure alteration (Chen et al., 2018; Qiu and Sheng, 2020). Spatially, WUE is highest in eastern China, followed by central China, while western China was the lowest in efficiency (Han et al., 2018). WUE of cities is not only related to their economic strength, but also connected to the water management level and the local physical endowment (Zhu and Tang, 2020). In general, the WUE is significantly lower in the agriculture sector than the industry and service sectors (Geng et al., 2019; Huang et al., 2021), higher in water-deficient or vulnerable regions than in water-rich regions (Guo et al., 2019), and lower in cities and towns than in rural areas (Hai et al., 2018).

In recent years, researchers have continuously renewed the evaluation methods for WUE. They have adopted a variety of methods including data envelopment analysis (DEA), stochastic frontier analysis (SFA) (Mu et al., 2016), the technique for order preference by similarity to ideal solution (TOPSIS) entropy weight fuzzy matter-element model (Liu et al., 2018), and the projection pursuit method (Fu et al., 2012), etc.. Among all these methodologies, DEA tends to be superior in evaluating multi-input and multi-output decision-making units (DMUs) based on a mathematical programming method, without dimensionless processing of data, prior identification of functional relationships, or setting non-subjective weights to each parameters (Wang et al., 2018a). DEA and its modified methodology gained extensive application worldwide, to comprehensively evaluate water use and assist decision-making (Hu et al., 2018; Gautam et al., 2020; Laureti et al., 2021). The national, provincial, and catchment research of WUE based on the methodology mentioned above reached broad consensus in China that the WUE is steadily improving, but still remains at a low level in general (Mu et al., 2016; Hai et al., 2018; Geng et al., 2019; Guo et al., 2019). Much more effort should be made to promote the construction of a water-saving society.

The WUE in Guangdong Province is slightly lower compared with the eastern metropolitan area (Sun and Liu, 2009; Zhang and Wu, 2020). The average WUE in Guangdong was only 26.9% of the national average in 2000, which increased to 31.9% in 2011. The utilization degree and reliability of water resources in Guangdong Province are restricted by the frequent droughts and floods, the limited water supply facilities, as well as the inadequate adjustment capacity. Apart from that, the widespread unreasonable water utilization, such as backward water resources management, the huge water resources waste, and improper sewage treatment, pose unprecedent threats on Guangdong water management (Chen, 2014). In addition, the spatial mismatch of the water resources and social-economic power inside Guangdong Province make the optimization of water utilization more challenging. The comprehensive development and utilization rate of water resources was 1.3% in Xijiang River Basin with plenty of water resources but a relatively weak economy, while in Dongjiang Basin where the water is in shortage and the economy is strong, 38.3% of water resources have been exploited. The per capita water availability is less than half of the national average in the core cities such as Guangzhou, Shenzhen, and Zhuhai. Three major practical demands were highlighted in the *Water* Conservancy and Water Resources Blue Book 2020: Guangdong-Hong Kong-Macao Greater Bay Area Water Resources Research Report: 1) solving the uneven spatial distribution of water resources in the Pearl River Basin, 2) optimizing the allocation of regional water resources development and utilization, and 3) guaranteeing the high-quality economic and social development and construction of the Guangdong–Hong Kong–Macao Greater Bay Area (Gu, 2002; SOHU, 2020).

Altogether, the current status of water resources restricts the sustainable development of Guangdong Province. Researchers and policy makers should comprehensively and systematically analyze the local WUE, clarify the temporal and spatial differentiation and its development trends, and explore strategies for the WUE optimization. Therefore, in this study we examined the WUE in prefectural cities in Guangdong Province during 2012–2018 based on the integrated Data Envelopment Analysis and Malmquist (DEA–Malmquist) model from spatial and temporal perspectives. We revealed the temporal and spatial characteristics of WUE changes in Guangdong Province and made predictions for the years 2019–2025. The study results will provide a theoretical basis and guidance for the guarantee of regional water resources.

METHODOLOGY

Study Area

Located in the southernmost region of the Chinese mainland, Guangdong Province has the most abundant light, heat, and water resources in China (**Figure 1A**). The region is dominated by a subtropical monsoon climate, with an annual average precipitation of 1,789 mm. Owing to the monsoon stability and tropical circulation, the precipitation has considerable temporal and spatial variability (SLT, 2020a). The average multiyear total water resources include 182 billion m^3 of surface water resources and 45 billion m^3 of groundwater resources with unbalanced temporal and spatial distributions. The topography of Guangdong is high in the north and low-lying in the south, resulting in severe flooding risk in the highly developed southern region, and low water storage in coastal terraces and low hilly areas. Due to the extensive non-point source pollution and ever-growing waste water effluence, the middle and lower reaches of Pearl River Delta are confronting with quality-induced water shortage. Guangdong Province is covered by Pearl River and its branches the Xijiang River, Dongjiang River, and Beijiang River systems (SLT, 2020b). Correspondingly, the province can be divided into four regions based on territory: the Pearl River Delta, eastern Guangdong, western Guangdong, and northern Guangdong (**Figure 1B**).

Research Methods

In this study, we employed the integrated DEA–Malmquist model to evaluate the WUE of Guangdong Province and each prefecture-level city from the years 2012–2018. The spatial/temporal variations and patterns of WUE were then analyzed using GIS spatial analysis and mathematical statistical analysis. The WUE prediction from 2019 until the end of the 14th Five-Year Plan period was carried out based on the annual growth rate estimation and integrated ARIMA and grey prediction model.

DEA–BCC Model

The DEA-CCR model was proposed in 1978 by American operational research experts Charness, Coopor, and Rhodes (CCR is the abbreviation of the developers) analyzing the relative efficiency during production process. DEA-CCR model assumes the production exhibits constant returns to scale, which means, for example the increase of 1% input will result in 1% production growth. This is clearly contrary to realism considering the scale effect of most industry and energy sectors. In the middle of 1980s, Banker, Charnes, and Cooper established a renovated DEA-BBC model, indicating that the production tends to show

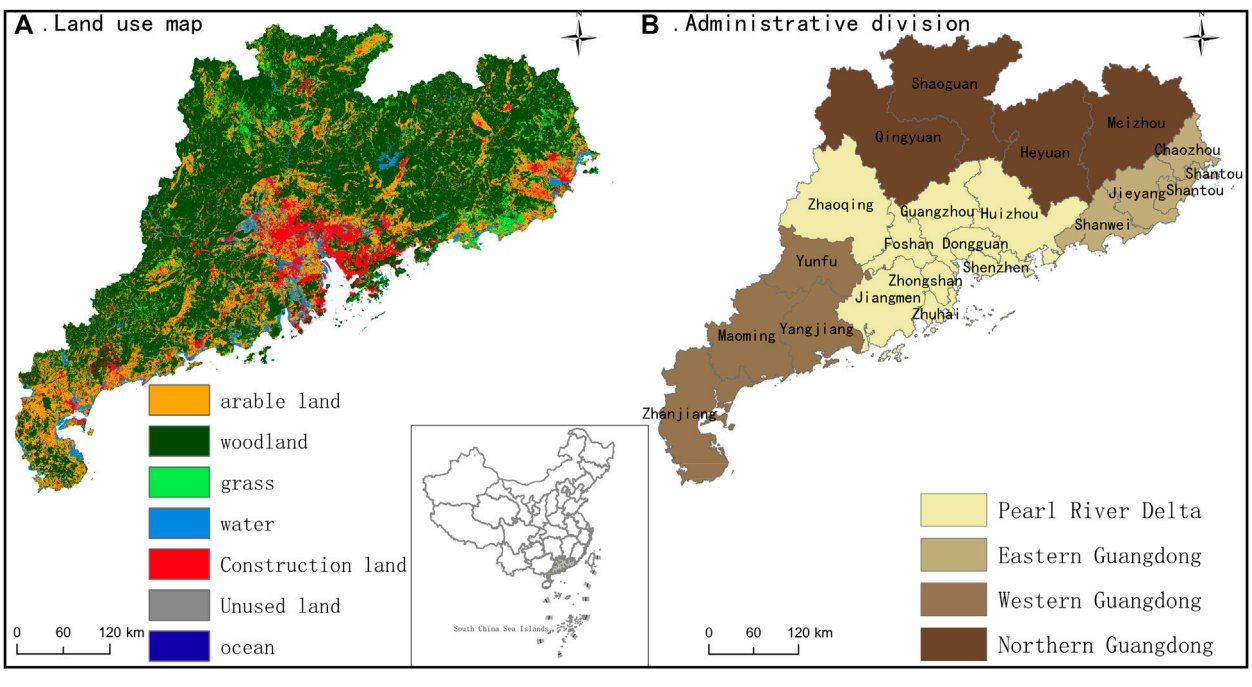


FIGURE 1 | Land use **(A)** and administrative division **(B)** map in Guangdong Province.

variable returns to scale during the estimation of the Decision-Making Units (DMUs) technology efficiency (Banker et al., 1984). The comprehensive efficiency can be correspondingly expressed by the multiplication of technical efficiency and scale efficiency (*crste = vrste*scale*). In addition, researchers can set the DEA-BBC model into input-oriented, output-oriented, or non-oriented according to the specific research objectives, where the orientation classifies the variable into optimized inputs, undesirable outputs, and overall outputs. Classically, the non-oriented model was more frequently adopted in environmental analysis, owing to its ability to deal with both desirable and undesired outputs simultaneously. DEA-BBC model has been widely applied in various professions, including water resources, energy industry, medical care, education, and banking (Castano and Cabanda, 2007; Sala-Garrido et al., 2012; Othman et al., 2016; Siampour et al., 2021; Sun et al., 2021), and was proven to be highly stable and reliable. As a result, we analyzed the WUE of Guangdong Province based on the DEA–BCC model in this study, trying to find out how to realize the high-efficiency utilization of water resources by reducing water resources and other component inputs, optimizing the desirable output and minimizing the negative output (Li and Cui, 2009). The model formula is shown below:

$$\begin{cases} min\left[\theta - \varepsilon\left(\sum_{i=1}^{m} sr^{-} + \sum_{r=1}^{n} si^{+}\right)\right] \\ \sum_{j=1}^{I} x_{ij}\lambda_j + si^{-} = \theta x_{ij0} \\ \sum_{j=1}^{I} y_{rj}\lambda_j - sr^{+} = y_{rj0} \\ \sum_{j=1}^{I} \lambda j = 1 \\ \lambda_j \geq 0, si^{-} \geq 0 \\ , sr^{+} \geq 0, j = 1, 2, 3, I, n \end{cases} \quad (1)$$

Where, x_{ij} refers to the *ith* input of the *jth* decision-making unit (DMU), y_{rj} means the *rth* output of the *jth* DMU, λj presents the planning decision variable, n is the DMU, ε demonstrates the non-Archimedean infinitesimal, and s_i^- and s_r^+ indicate the slack variable vectors. The effective value of the DMU is θ.

Generally, the DMU is considered to be DEA-efficient when $\theta = 1$, and $s_i^- = s_r^+ = 0$. In such a situation, the DMU is optimal regardless of input or output perspective, and the production process reaches technical and scale efficiency. If $\theta = 1$, and one of the slack variable vectors is greater that 0, then the DMU is considered to be weakly efficient, the production process is either technically inefficient or scale inefficient. If $\theta < 1$, then the DMU is DEA-inefficient, neither technical efficiency nor scale efficiency is realized.

Malmquist Index Model

The Malmquist index model was developed based on the DEA model, which can reflect the dynamic change of the DMU efficiency by estimating the efficiency of input and output using the distance function. This model usually assumes that input elements remain unchanged and the output need to be optimized. Distance functions in different periods are established to represent the corresponding temporal efficiency, and each decision unit is compared with the optimized front plane to construct the dynamic model of productivity.

The Malmquist index can be decomposed into technological efficiency change (*effch*) and technological progress (*techch*) indices. The *effch* can be further divided into pure efficiency change (*pech*) and scale efficiency change (*sech*) indices. These decompositions indicate that the productivity growth is for one thing caused by the development of science and technology (*techch*), and for another comes from the organization and management progress (*effch*), including the pure efficiency change (*pech*) and scale effect improvement (*sech*). Thus, *Effch* represents the level of organization and management and the ability to comprehensively allocate resources across the region. The *techch* refers to the development and progress of science and technology related to the development and use of water resources. *Pech* primarily means the production efficiency of enterprises affected by factors such as management and technology, and the *sech* is the efficiency change caused by the expansion or reduction of production scale.

The indices were decomposed as follows:

$$M\left(x^{t+1}, y^{t+1}, x^{t}, y^{t}\right) = \sqrt{\left[\frac{D^{t}\left(x^{t+1}, y^{t+1}\right)}{D^{t}\left(x^{t}, y^{t}\right)} \times \frac{D^{t+1}\left(x^{t+1}, y^{t+1}\right)}{D^{t+1}\left(x^{t}, y^{t}\right)}\right]} \times \frac{D^{t}\left(x^{t+1}, y^{t+1}\right)}{D^{t}\left(x^{t}, y^{t}\right)} \quad (2)$$

$$Techch = \sqrt{\left[\frac{D^{t}\left(x^{t+1}, y^{t+1}\right)}{D^{t}\left(x^{t}, y^{t}\right)} \times \frac{D^{t+1}\left(x^{t+1}, y^{t+1}\right)}{D^{t+1}\left(x^{t}, y^{t}\right)}\right]} \quad (3)$$

$$Effch = Pech \times Sech = \frac{D^{t}\left(x^{t+1}, y^{t+1}\right)}{D^{t}\left(x^{t}, y^{t}\right)} \quad (4)$$

$$Tfpch = \left(Effch \times Sech\right)Tech \quad (5)$$

where $\left(x^{t+1}, y^{t+1}\right), D^{t}\left(x^{t}, y^{t}\right)$ are output distance functions; $M\left(x^{t+1}, y^{t+1}, x^{t}, y^{t}\right)$ represents the productivity index in the period t and $t + 1$; and *tfpch* means the aggregate productivity index. *tfpch* > 1, represents the growth in efficiency or productivity, *tfpch* = 1 indicates unchanged efficiency or productivity, while tfpch <1 reflects decline in efficiency or productivity.

Data Sources

The data we used in this study were primarily collected from the Guangdong Statistical Yearbook 2012–2020, the Water Resources Bulletin of Guangdong Province 2012–2020, and the statistical yearbooks of various prefecture-level cities in Guangdong Province of the corresponding year.

Index Selection

DEA evaluates the decision-making problem of water resource utilization with multiple inputs and outputs, showing high flexibility and simplicity, irrespective of the correlation between variables, the inefficiency distribution, and the type of production process (Nataraja and Johnson, 2011). It is not necessary to set a specific production function or to specify

TABLE 1 | Output-input indices of water-use efficiency.

Index type	Index	Specific index
Input	Labor input	Number of employees (10,000)
	Water resource input	Industrial water consumption (100 million m^3)
		Domestic water consumption for residents (100 million m^3)
		Eco-environmental water consumption (100 million m^3)
Output	Economic benefit output	GDP (100 million yuan)
	Wastewater output	Quantity of wastewater effluent (100 million tons)

TABLE 2 | Prediction methods and sources of index data.

Index type	Index	Prediction method	Basis
Input	Industrial water consumption (100 million m^3)	Grey prediction model	Wu et al. (2017)
	Domestic water consumption for residents (100 million m^3)		
	eco-environmental water consumption (100 million m^3)		
	Number of employees (10,000)	ARIMA model	Alho (2004)
Output	GDP (100 million yuan)		
	Quantity of wastewater effluent (100 million tons)	Grey prediction model	Li et al. (2017)

the distribution of the error term. It is worth mentioning that the ratio cannot serve as an input or output indicator, meanwhile the selected indicators should be authentic reflecting a basic production relationship. According to the connotation of WUE and the data accessibility of Guangdong, combined with the relevant studies, we took labor and water resource as input variables, and the economic benefit and wastewater as output index (**Table 1**) (Hu et al., 2019; Wang et al., 2021). In terms of water resource input, we primarily focus on the industrial water use, domestic water for residents and eco-environmental water use, considering the highly developed economy, densely concentrated population, and the environment-friendly development orientation of Guangdong province.

Index Prediction

Population growth and socio-economic development in Guangdong Province are supposed to maintain at high rates during the 14th Five Year Plan Period (2021–2025), which will inevitably intensify the conflicts between water supply and demand among different regions and sectors. It is thus essential to forecast the WUE of the corresponding period. We estimated the various indices of Guangdong Province in 2019–2025 based on ARIMA and the grey prediction model, and then evaluated WUE in Guangdong Province for 2019–2025 using the DEA–Malmquist model (**Table 2**).

The flow chart of this research is shown as **Figure 2**.

RESULTS AND DISCUSSION

Temporal Variation of DEA Efficiency in Guangdong Province

DEA efficiency mainly represents the overall WUE of the entire region. According to the results of the DEA–Malmquist model, the annual overall WUE in Guangdong Province was on an increasing trend during the study period, from 0.896 in 2012 to 0.936 in 2018 (as shown in **Figure 3**). In terms of subregion, the WUE in the Pearl River Delta was the highest, followed by that in northern Guangdong, and eastern Guangdong, while western Guangdong showed the lowest in WUE. Apart from spatial variation, the WUE significantly fluctuated over time in different regions. The WUE in the Pearl River Delta kept to around 0.9, significantly higher than that in other regions during the whole study period. In northern Guangdong, the WUE was increased gradually to 0.943 in 2016 and then declined by nearly 9% in the following 2 years. The WUE in eastern Guangdong was relatively low and showed large fluctuations, while in western Guangdong the averaged WUE was 0.712, much lower than other sub regions. In perspective of the decomposed efficiency, all the regions reached technological efficiency except western Guangdong. Pearl River Delta and eastern Guangdong

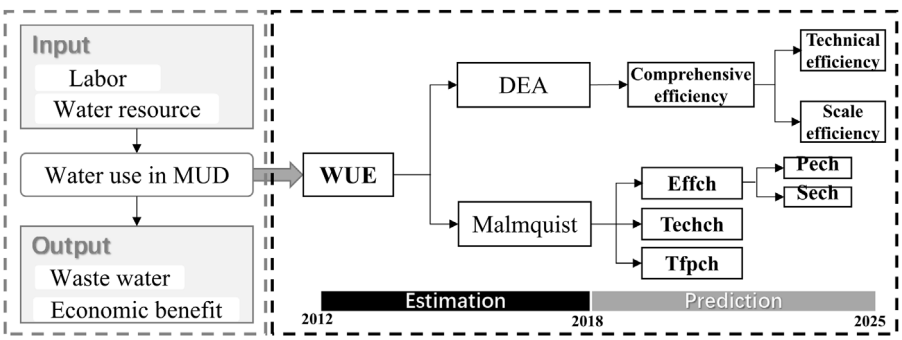


FIGURE 2 | The flowchart of this study.

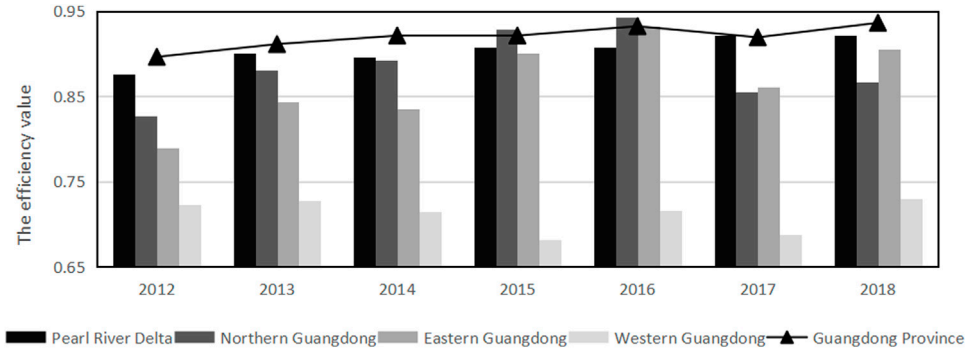


FIGURE 3 | Average WUE in different regions of Guangdong Province.

exhibited the highest and lowest scale efficiency respectively (**Table 3**).

According to the Malmquist decomposition model, the annual average water resources total factor productivity change (*tfpch* 1.005), the *effch* (1.008), *pech* (1.004), and *sech* (1.004) values in Guangdong Province were all larger than 1, while the annual averaged *techch* (0.997) were lower than 1 (**Table 4**). These results demonstrated that the overall growth of the WUE in Guangdong Province benefitted mainly from the promotion of the comprehensive ability of enterprise organization and management (*Effch*). Besides, the optimization of regional water allocation ability (*pech*) and expansion of scale (*sech*) made accessory contributions to the increasing WUE. In contrast, although the new technologies of water resource conservation and high efficient utilization were widely adopted, the technical limitations for the water resources utilization, indicated by the low *techch*, were still obvious in Guangdong. Currently, traditional manufacture of extensive type, represented by high water consumption, high water contamination, and high waste water emission still occupies a large proportion in Guangdong. Introduction and implementation of superior technology promoting industrial water resource saving, circulation, and pollution control are strongly recommended to improve the local WUE.

WUE Growth in Typical Years

2016–2017 was a typical period when the WUE growth was driven by strong technology progress, while in other stages the WUE increases were mainly conducive to scale expansion or management advantage. In 2016–2017, the water conservancy investment of Guangdong Province reached 31 billion yuan, creating a new record. Furthermore, the *Provincial Water Resources Conservation and Protection Special Fund* initiated by the Water Resources Department strongly promotes investment in water related research and development. Financial support and scientific research progress contributed to the construction of water conservancy facilities and the upgradation of water-saving technologies in Guangdong Province and promoted the *techch* to a large extent, thereby the improvement of WUE was achieved. As shown in **Table 4**, *effch*, *pech*, and sech values in 2016–2017 were all less than 1, which indicates that Guangdong Province made a deficient effort in the resource scale allocation, management methods, and management organization structure compared with scientific investment in this period. In 2018, the *effch* and other decomposition indices increased to greater than 1. The policies issued in 2018, Guangdong Province Urban Management Regulations (2018 Revision) and Strictest Water Resources Management Assessment Measures Implemented in Guangdong Province, greatly promoted the optimization of resource allocation,

TABLE 3 | Averaged decomposed WUE of various regions in Guangdong Province from 2012 to 2018.

Region	Comprehensive efficiency (crste= vrste* scale)	Technical efficiency (vrste)	Scale efficiency (scale)
Guangdong	0.919	0.953	0.966
Pearl River Delta	0.904	1	0.904
Eastern Guangdong	0.867	1	0.867
Western Guangdong	0.712	0.808	0.882
Northern Guangdong	0.885	1	0.885

TABLE 4 | Malmquist Index and its Decomposition Results in Guangdong Province from 2012 to 2018.

Year	*Effch*	*Effch = Pech * Sech*		*Techch*	*Tfpch*
		Pech	*Sech*		
2012–2013	**1.020 (+)**	**1.011 (+)**	**1.009 (+)**	0.985	**1.005 (+)**
2013–2014	**1.011 (+)**	0.998	**1.013 (+)**	0.988	0.998
2014–2015	1	**1.009 (+)**	0.991	0.994	0.993
2015–2016	**1.014 (+)**	**1.004 (+)**	**1.009 (+)**	0.998	**1.011 (+)**
2016–2017	0.986	0.994	0.991	**1.028 (+)**	**1.013 (+)**
2017–2018	**1.017 (+)**	**1.006 (+)**	**1.011 (+)**	0.991	**1.008 (+)**
Average value	**1.008 (+)**	**1.004** (+)	**1.004** (+)	0.997	**1.005** (+)

The bold value represent the increasing trend.

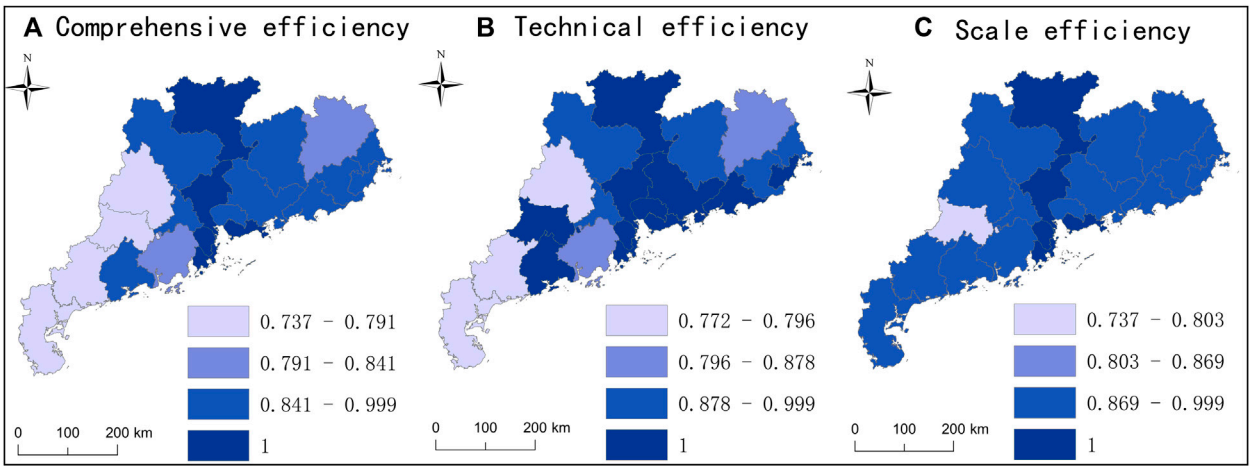


FIGURE 4 | WUE of the prefecture-level cities in Guangdong Province in 2012–2018.

management strategy, and organization structure in Guangdong Province. In contrast, the *techch* value in 2017–2018 was less than 1, which meant the technological innovation was relatively less propulsive. According to the analysis of the typical periods from 2016 to 2018, it is clear that the technical innovation seems be harder to achieve in the annual slice, in contrast to the gradual management or scale effect, but the technical innovation can promote the WUE to a larger extent once it works.

Spatial Pattern of WUE

As shown in **Figure 4**, only five cities (Guangzhou, Shenzhen, Zhuhai, Zhongshan, and Shaoguan) reached the DEA efficiency in water utilization. As the leading forces of economic development in Guangdong Province, the Guangzhou, Shenzhen, Zhuhai, and Zhongshan were superior to other cities in terms of technological innovation and industrial structure configuration, and their average WUE was higher than surrounding regions. The water resources in western Guangdong were relatively abundant, but the average WUE was low because of many technical developments and physical restrictions.

From the Malmquist index and its decomposition analysis, about half of the prefecture-level cities had made progress in the WUE during the study period. Among them, The WUE in Jieyang increased with the highest rate in the technological innovation, organizational management, water resource allocation, and scale expansion aspects. In addition, Jieyang was the only city which achieved the comprehensive progress. Despite the organizational management and scale effect of Guangzhou, Shenzhen and Shantou remained at a stable level, the remarkable technological innovation of the three cities promoted the WUE growth, especially Shenzhen, representing the most rapid technological advancement. The WUE increase of Foshan, Maoming, and Shanwei were integrally driven by technological progress and scale expansion. Chaozhou gained momentum from scale expansion, which helped the WUE to continue ascending. Excellent enterprise management, reasonable resources allocation, and stable technical development drove the WUE growth in Meizhou, Zhanjiang, and Qingyuan. Apart from the above mentioned cities, the WUEs of the rest of the regions were decreasing, mainly restricted by the technological degradation, except Zhaoqing, which is hindered by a combination of technological and organizational weakness (**Table 5**).

In terms of subregion, the WUE growth of PRD and northern Guangdong were mainly dominated by scale expansion, while eastern Guangdong realized the comprehensive progress of the highest WUE increase due to a strong scale effect, slight technological progress, and stable organizational management. The rising WUE of western Guangdong was driven by similar scale and technological progress, despite the relative deficiency in enterprise management and resources allocation (**Table 5**).

In general, the technology innovation is the dominant factor controlling the water use efficiency and productivity. Nine and six prefectural cities improved and regressed their water use productivity, respectively, due to the technological parameters. Furthermore, the variation of prefecture city WUE showed strong spatial heterogeneity and spatial agglomeration effect in the study period. For example, the technological regression region was mainly concentrated in middle-east and middle-west region (**Figure 5**). In the sub region perspective, the Pearl River Delta and northern Guangdong regions were advantageous in the scale expansion effect, the Western Guangdong region had scale and technological preponderance, while the eastern Guangdong region made progress in all-round parameters.

Prediction on WUE

DEA Prediction

The predictions showed that the WUE in Guangdong Province will continue to improve over time. The average overall efficiency in 2019–2025 would reach 0.933, around 2% higher than the

TABLE 5 | Malmquist index and its decomposition of average water resources in prefecture-level cities and regions in the last 9 years.

City	*Effch*	*Effch = pech [a]sech*		*Techch*	*Tfpch*	Type
		Pech	Sech			
Jieyang	1.031 (+)	1.013 (+)	1.018 (+)	1.022 (+)	1.054 (+)	Comprehensive progress
Guangzhou	1	1	1	1.001 (+)	1.001 (+)	Technological innovation
Shenzhen	1	1	1	1.051 (+)	1.051 (+)	
Shantou	1	1	1	1.022 (+)	1.022 (+)	
Foshan	1.006 (+)	1	1.006 (+)	1.029 (+)	1.036 (+)	
Shanwei	1.008 (+)	1	1.008 (+)	1.013 (+)	1.021 (+)	
Maoming	0.985	0.983	1.002 (+)	1.025 (+)	1.01 (+)	
Chaozhou	1.041 (+)	1.018 (+)	1.023 (+)	0.973	1.013 (+)	Scale advantage
Meizhou	1.024 (+)	1.025 (+)	0.999	0.981	1.005 (+)	Organizational management advantage
Zhanjiang	1.039 (+)	1.041 (+)	0.998	1.016 (+)	1.055 (+)	
Qingyuan	1.005 (+)	1.006 (+)	0.998	1.004 (+)	1.009 (+)	
Zhuhai	1	1	1	0.975	0.975	Technological regression
Huizhou	1	1	1	0.984	0.984	
Shaoguan	1	1	1	0.972	0.972	
Heyuan	1.002 (+)	1	1.002 (+)	0.992	0.993	
Dongguan	1.012 (+)	1	1.012 (+)	0.986	0.998	
Zhongshan	1	1	1	0.961	0.961	
Jiangmen	1.005 (+)	1.01 (+)	0.995	0.972	0.977	
Yangjiang	1	1	1	0.993	0.993	
Yunfu	1.007 (+)	1	1.007 (+)	0.989	0.995	
Zhaoqing	0.999	0.984	1.015 (+)	0.984	0.983	Comprehensive restrictions
Guangdong	**1.008** (+)	**1.004** (+)	**1.004** (+)	**0.997**	**1.005** (+)	**—**
PRD[a]	**1.008 (+)**	**1**	**1.008 (+)**	**0.998**	**1.006 (+)**	**Scale progress**
EG[a]	**1.008 (+)**	**1**	**1.008 (+)**	**0.98**	**0.988**	
WG[a]	**1.023 (+)**	**1**	**1.023 (+)**	**1.001 (+)**	**1.024 (+)**	**Comprehensive progress**
NG[a]	**1.002 (+)**	**0.989**	**1.012 (+)**	**1.013 (+)**	**1.015 (+)**	**Scale and technology progress**

The bold typing means the four regions and the province, while the others are the prefecture-level cities.
[a]PRD represent Pearl River Delta, while EG, WG and NG are the abbreviation of eastern, western and northern Guangdong.

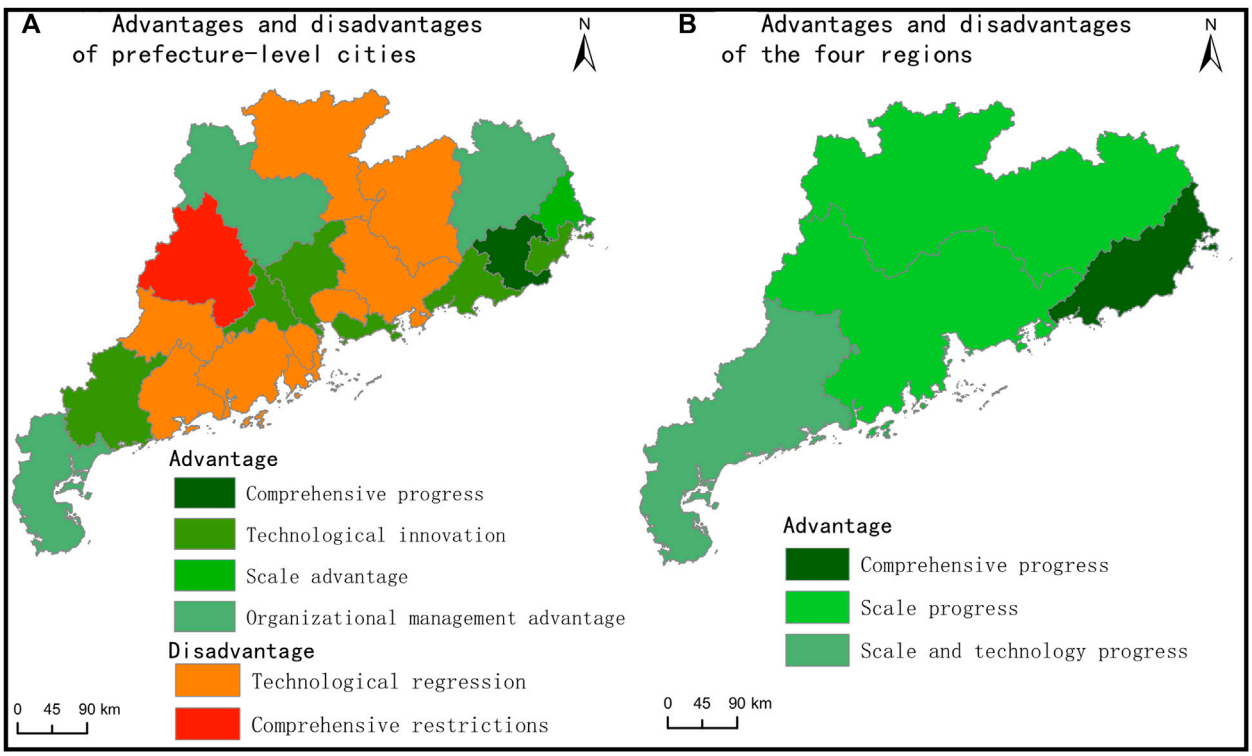


FIGURE 5 | Malmquist index and decomposition trend map of water resource utilization in prefecture-level cities in Guangdong Province in the last 7 years.

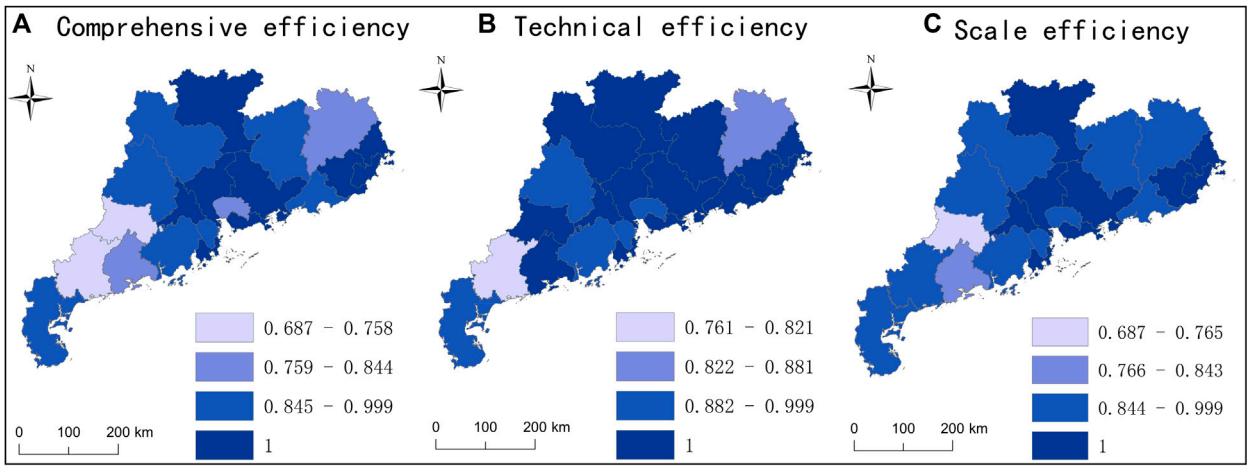


FIGURE 6 | Efficiency distribution map in prefecture-level cities in 2019–2025.

2012–2018 level, although still falling behind DEA efficiency. With the continuous emphasis on science and technology investment, the technological efficiency would be largely improved. However, the scale efficiency would keep constant or grow slowly limited to the scale benefit decline principle. Spatially, the regions where WUE reached DEA efficiency would extend from the Pearl River Delta to the adjacent western and northern Guangdong regions during 2019–2025. The spatial heterogeneity tends to weaken with the promotion of the WUE (**Figure 6**).

Malmquist Prediction

As shown in **Table 6**, the water resource utilization productivity value was greater than 1, and the average annual growth rate was maintained at 0.4–0.5%, indicating the improvement of the total factor efficiency during the prediction period. On one hand, *Techch* and *Pech* serve as the main and secondary driving force for improving efficiency respectively, with the correlation coefficient between technological progress and total factor productivity of water resources reaching 0.863. On the other hand, the *sech* would improve at the beginning and then enter the scale benefit decline stage, which would restrict the upgrading of productivity. Guangdong Province would make progress in the implementation of water resource saving, protection, and recycle utilization technology. The regional comprehensive resource allocation and enterprise management capabilities are expected to be strengthened.

In terms of spatial pattern, during the 14th Five-Year Plan period, rapid comprehensive growth of the productivity and WUE in Dongguan, Zhanjiang, and Heyuan would be witnessed. Coastal cities including Shenzhen, and Foshan would achieve innovative growth conducive to the development of science and technology. Traditional high water efficiency cities, Guangzhou, Shaoguan, and Zhuhai are threatened by the relatively declined technology innovation and the potential retrogress of the productivity and WUE. Although the organizational management, allocation capability and scale benefit tends to expansion in Jiangmen city, the low technology investment might hinder the local productivity and WUE. All the other cities would suffer a slight decline of overall productivity and WUE, owing to the combined weakness of the resource allocation, technology innovation, enterprise management, or shrinking scale return (**Table 7** and **Figure 7**).

TABLE 6 | Malmquist index and its decomposition results in Guangdong Province in 2019–2025.

Year	*Effch*	*Effch = pech *sech*		*Techch*	*Tfpch*
		Pech	*Sech*		
2019–2020	1.006	1.003	1.003	0.995	1
2020–2021	1.006	1.003	1.003	1.002	1.008
2021–2022	0.999	1	1	1.009	1.009
2022–2023	1.001	1.001	1.001	0.997	0.998
2023–2024	1.002	1.004	0.998	1.013	1.015
2024–2025	0.995	0.999	0.996	1.008	1.004
Average	1	1	1	1.004	1.004

Significance of Parameters

We conducted correlation and automatic linear modeling regression to analyze the significance of the input and output parameters to the efficiency. The correlation results showed that the output parameter, GDP, and waste water discharge, were correlated with efficiency relatively higher compared to the input parameters. The increasing GDP and its investment to management renewal and technology upgrading promoted the growth of WUE (**Table 8**).

As shown in **Figure 8**, GDP, the amount of employees and the waste water discharge are the main influencing factors to the WUE. Specifically, scale efficiency was mainly controlled by GDP, while the amount of employees was the strongest influence to the technical and comprehensive efficiency. Waste water discharge and GDP played the secondary role in the technical and comprehensive WUE. New technical approach reducing the employee investment and

TABLE 7 | Average Malmquist index and its decomposition value of cities in Guangdong Province in 2019–2025.

City	*Effch*	*Effch = pech *sech*		*Techch*	*Tfpch*	Type
		Pech	*Sech*			
Dong guan	1.044	1.011	1.033	1.004	1.048	Comprehensive growth
Zhanjiang	1.019	1.019	1.001	1.038	1.058	
Heyuan	1.003	1	1.003	1.04	1.043	
Shenzhen	1	1	1	1.039	1.039	Innovative growth
Shantou	1	1	1	1.02	1.02	
Foshan	1	1	1	1.059	1.059	
Huizhou	1	1	1	1.006	1.006	
Chaozhou	1	1	1	1.013	1.013	
Jieyang	1	1	1	1.066	1.066	
Maoming	0.999	1	0.999	1.003	1.001	
Guangzhou	1	1	1	0.989	0.989	Innovative decline
Shaoguan	1	1	1	0.965	0.965	
Zhuhai	1	1	1	0.983	0.983	
Jiangmen	1.022	1.006	1.016	0.977	0.998	
Zhaoqing	1.021	1.021	0.999	0.97	0.99	Comprehensive degradation
Shanwei	0.974	1	0.974	0.999	0.973	
Yangjiang	0.992	1	0.992	1.002	0.994	
Qingyuan	0.998	1	0.998	0.996	0.993	
Yunfu	0.981	1	0.981	0.991	0.972	
Meizhou	0.996	0.986	1.01	0.987	0.984	
Zhongshan	0.988	0.993	0.995	0.946	0.935	
Guangdong	**1.002**	**1.002**	**1**	**1.004**	**1.006**	—

The bold represents the general condition of Guangdong Province.

waste water output would highly improve the technical and comprehensive efficiency in future endeavors. Maintaining the GDP growth under the same or lower water consumption would stimulate the WUE in all perspectives.

Management Implications

Substantial efforts toward water resource protection and water ecological/environmental control have been made on the national level, which improves the WUE overall. High developed metropolises like Jingjintang (JJT), Yangtze River Delta (YRD) and Peral River Delta (PRD) depict higher WUE index, owing to the more developed water allocation strategy and water utilizing technique. The WUE in Guangdong province is higher than the YRD (Shanghai and Jiangsu), but lower than the JJT region (Liu et al., 2020; Shi et al., 2020). Although some core PRD cities, such as Guangzhou, showed the best decoupling status between the urban economic output and water usage (Wang et al., 2018b), Guangdong Province is still enconfronting with low efficiency problems especially in industry water use. As the one of the leading industry provinces, waste water discharge of Guangdong Province ranks first among all the Chinese Provinces, strongly restricting the further improvement of WUE. Professional disposal techniques are in urgent need to effectively handle the huge wastewater discharge containing nitrogen and other harmful chemicals. Research and development investment in waste water treatment infrastructure and enhancement of wastewater treatment capacity are critical for WUE

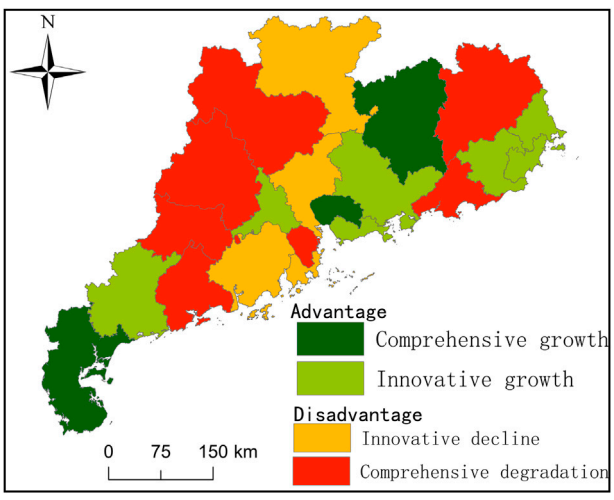


FIGURE 7 | Development of predicted WUE in Guangdong Province during 2019–2025.

TABLE 8 | Average Malmquist index and its decomposition value of cities in Guangdong Province in 2019–2025.

Correlation	GDP	WW	IWC	RWC	EWC	Employee
Scale efficiency	**0.250****	**0.212****	0.074	0.105	−0.051	0.081
Technical efficiency	**0.189***	**0.216****	0.170*	0.052	0.173*	0.077
Comprehensive efficiency	**0.208***	**0.264****	0.145	0.013	−0.084	−0.018

*Note, ** and * refer to the correlations which passed the 0.01 and 0.05 significance test. WW represents the waste water discharge. IWC, RWC, and EWC mean the industrial, residential, and eco-environmental water consumption individually. The bold values indicate the higher correlation of GDP and WW with other parameters.*

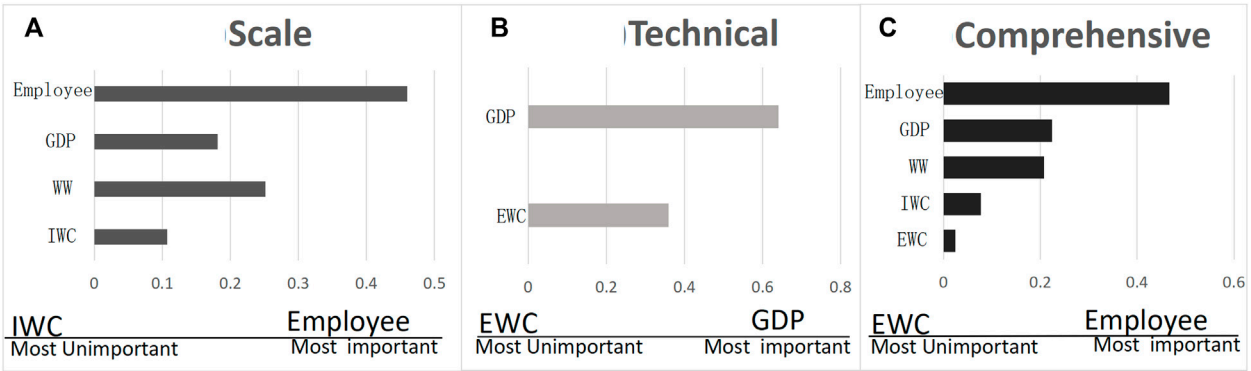


FIGURE 8 | Significance of each parameter.

improvement. In addition, the regional disparities in industrial WUE make new demands on reliable and effective water allocation projects and up-down stream linkage water dispatch. With respect to urban and rural water use, adjusting supply structure, adopting discharge-reducing techniques, and reclaiming domestic waste water might be effective to promote the WUE.

CONCLUSION

Water is a significant resource supporting socio-economic growth and maintaining environmental health. The water utilization efficiency (WUE) reflected the intensity and potential of water utilization and development. Moreover, improving WUE is conducive to the sustainable development of the economic–social–ecological system on a regional and global scale. In this paper, we evaluate and predict the water utilization efficiency (WUE) in Guangdong Province based on the DEA–Malmquist model. The results revealed the gradual increase of the WUE, from 0.896 in 2012 to 0.936 in 2018. The WUE growth was for one thing stemmed from the scientific and technological progress, and for another driven by the improvement of regional resource allocation capability and the scale expansion benefit. The comprehensive promotion of the enterprise organization and management serves as an important factor affecting the total factor productivity of water resources. In terms of spatial distribution, the WUE in Pearl River Delta was higher than that of northern Guangdong and eastern Guangdong, while western Guangdong was lowest in WUE. Pearl River Delta made WUE progress due to the advantages of organization and management, as well as the scale benefit. WUE in eastern and western Guangdong were basically effective due to the integrated advantages of scale benefit, technological innovation, or organizational management. Northern Guangdong is relatively deficient in water resources, technological innovation, and the allocation of various resources. The spatial heterogeneity of WUE was on decline trend. Prediction analysis suggest that the technological innovation and the integrated management capability in Guangdong Province will be continuously promoted during the 14th Five Year Plan period, while the scale expansion benefit tends to diminish. The WUE and productivity will be comprehensively improved, and the spatial difference will be continually narrowed along time. This study responds to improve the WUE, to alleviate the water resource pressure, and to achieve optimal water use under the rapid regional socioeconomic development. The results provide important perspectives to policy-makers on how to balance the economic growth, the ecological health, and the human-water system, to ultimately pursue the sustainable development.

DATA AVAILABILITY STATEMENT

The original contribution presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

SS conceptualized the research and managed the funding; LC was responsible for the software, the calculation and organized the original manuscript; YX contributed in data collection and processing.

FUNDING

This research is supported by the Guangdong Province Universities and Colleges Pearl River Scholar Funded Scheme (2019), Natural Science Foundation of Guangdong Province (2020A1515011065) and Key Special Project for Introduced Talents Team of Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou) (GML2019ZD0301).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fenvs.2022.819693/full#supplementary-material

REFERENCES

Adeel, Z. (2017). A Renewed Focus on Water Security within the 2030 Agenda for Sustainable Development. *Sustain. Sci.* 12, 891–894. doi:10.1007/s11625-017-0476-7

Alho, J. M. (2004). Population Forecasting Theory, Methods and Assessments of Accuracy. The Magnitude of Error Due to Different Vital Processes in Population Forecasts. *Int. J. Forecast.* 8, 301–314. doi:10.1016/0169-2070(92)90049-f

Banker, R. D., Charnes, A., and Charnes, W. W. (1984). Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis. *Manage. Sci.* 30, 1078–1092. doi:10.1287/mnsc.30.9.1078

Castano, M. C. N., and Cabanda, E. C. (2007). Performance Evaluation of the Efficiency of Philippine Private Higher Educational Institutions: Application of Frontier Approaches. *Int. Trans. Oper. Res.* 14, 431–444. doi:10.1111/j.1475-3995.2007.00599.x

Chen, T. L. (2014). Changes, Problems and Suggestions of Water Resources in Guangdong Province. *People's Tribune* (34), 222–223. doi:10.16619/j.cnki.rmlt.2014.34.003 (in Chinese).

Chen, W., Du, J., and Chang, J. J. (2018). Utilization Efficiency of Water Resources in Wuhan Urban Agglomeration. *Resour. Environ. Yangtze Basin* 27 (06), 1251–1258. doi:10.11870/cjlyzyyhj201806008 (in Chinese).

Chen, X. X., Zhuang, Y. N., Liu, L., and Wang, L. N. (2020). Water Use Efficiency in Shanxi and Shanxi Province Ananlyzed Using the Super Efficiency DEA Model. *J. Irrigation Drainage* 39 (10), 138–144. doi:10.13522/j.cnki.ggps.2020029 (in Chinese).

Ding, X., Fu, Z., and Jia, H. (2019). Study on Urbanization Level, Urban Primacy and Industrial Water Utilization Efficiency in the Yangtze River Economic Belt. *Sustainability* 11, 6571. doi:10.3390/su11236571

Fan, X. Z., Chen, H. Z., Zhang, P. N., Chen, H., and Zhang, F. N. (2008). Investigation Report on Water Energy Resources Development and Utilization Management in USA and Canada. *China Water Power & Electrification* (04), 23–28+36. doi:10.3969/j.issn.1673-8241.2008.04.009 (in Chinese).

Fu, Q., Jiang, Q., and Wang, Z. (2012). "Sustainability Assessment of Regional Water Resources Use Based on PSO-PPE," in *Computer and Computing Technologies in Agriculture V*. Editors D. Li and Y. Chen (Beijing: Springer), 383–390. doi:10.1007/978-3-642-27275-2_43

García-Valiñas, M. A., and Muñiz, M. A. (2007). Is DEA Useful in the Regulation of Water Utilities? A Dynamic Efficiency Evaluation (A Dynamic Efficiency Evaluation of Water Utilities). *Appl. Econ.* 39 (2), 245–252. doi:10.1080/00036840500428054

Gautam, T. K., Paudel, K. P., and Guidry, K. M. (2020). An Evaluation of Irrigation Water Use Efficiency in Crop Production Using a Data Envelopment Analysis Approach: A Case of Louisiana, USA. *Water* 12 (11), 31–93. doi:10.3390/w12113193

Geng, Q., Ren, Q., Nolan, R. H., Wu, P., and Yu, Q. (2019). Assessing China's Agricultural Water Use Efficiency in a green-blue Water Perspective: A Study Based on Data Envelopment Analysis. *Ecol. Indicators* 96, 329–335. doi:10.1016/j.ecolind.2018.09.011

Gu, J. H. (2002). The Causes and Countermeasures of the Prominent Contradiction between Supply and Demand of Water Resources in Guangdong Province. *Pearl River* (1), 17–20. doi:10.3969/j.issn.1001-9235.2002.01.006 (in Chinese).

Guo, S.-D., Li, H., Zhao, R., and Zhou, X. (2019). Industrial Environmental Efficiency Assessment for China's Western Regions by Using a SBM-Based DEA. *Environ. Sci. Pollut. Res.* 26, 27542–27550. doi:10.1007/s11356-019-06062-5

Hai, X., Li, W. F., and Han, L. J. (2018). Study on the Domestic Water Use Efficiency between Urban and Rural Areas of Beijing-Tianjin-Hebei Urban Agglomeration, China. *J. Water Resour. Water Eng.* 29 (02), 27–33. doi:10.11705/j.issn.1672-643X.2018.02.05 (in Chinese).

Han, W. Y., Chen, X. P., Zhang, Z. L., Wang, N. F., and Yu, Y. H. (2018). Analysis on Spatiotemporal Structure of Water Resources Utilization Efficiency of the Orefecture Level Cities and beyond in China. *Res. Soil Water Conservation* 25 (02), 354–360. doi:10.13869/j.cnki.rswc.2018.02.051 (in Chinese).

Hu, M., Hu, Y., and Yuan, J. (2019). Benchmarking on Water Resource Utilization Efficiency of Prefecture-Level Cities in Jiangxi, China: A Bootstrap-DEA Approach with Three-Stage DEA Models. *Environ. Nat. Resour. J.* 9, 14. doi:10.5539/enrr.v9n3p14

Hu, Z., Yan, S., Yao, L., and Moudi, M. (2018). Efficiency Evaluation with Feedback for Regional Water Use and Wastewater Treatment. *J. Hydrol.* 562, 703–711. doi:10.1016/j.jhydrol.2018.05.032

Huang, Y., Huang, X., Xie, M., Cheng, W., and Shu, Q. (2021). A Study on the Effects of Regional Differences on Agricultural Water Resource Utilization Efficiency Using Super-Efficiency SBM Model. *Sci. Rep.* 11, 9953. doi:10.1038/s41598-021-89293-2

Khaskheli, M. A., Abro, M. I., Chand, R., Elahi, E., Khokhar, F. M., Majidano, A. A., et al. (2021). Evaluating the Effectiveness of Eggshells to Remove Heavy Metals from Wastewater. *Desalination Water Treat* 216, 239–245. doi:10.5004/dwt.2021.26807

Laureti, T., Benedetti, I., and Branca, G. (2021). Water Use Efficiency and Public Goods Conservation: A Spatial Stochastic Frontier Model Applied to Irrigation in Southern Italy. *Socio-Economic Plann. Sci.* 73, 100856. doi:10.1016/j.seps.2020.100856

Li, S., Meng, W., and Xie, Y. (2017). Forecasting the Amount of Waste-Sewage Water Discharged into the Yangtze River Basin Based on the Optimal Fractional Order Grey Model. *Int. J. Environ. Res. Public Health* 15, 20. doi:10.3390/ijerph15010020

Li, X., and Cui, J. (2009). "Real-Time Water Resources Allocation: Methodology and Mechanism," *in 2009 IEEE International Conference on Industrial Engineering and Engineering Management*, Beijing, China, December 8–11, 2009, 1637–1641. doi:10.1109/ieem.2009.5373121

Liu, K.-D., Yang, G.-L., and Yang, D.-G. (2020). Investigating Industrial Water-Use Efficiency in mainland China: An Improved SBM-DEA Model. *J. Environ. Manage.* 270, 110859. doi:10.1016/j.jenvman.2020.110859

Liu, X., Qi, Y., Li, F., Yang, Q., and Yu, L. (2018). Impacts of Regulated Deficit Irrigation on Yield, Quality and Water Use Efficiency of Arabica Coffee under Different Shading Levels in Dry and Hot Regions of Southwest China. *Agric. Water Manage.* 204, 292–300. doi:10.1016/j.agwat.2018.04.024

Mu, L., Fang, L., Wang, H., Chen, L., Yang, Y., Qu, X. J., et al. (2016). Exploring Northwest China's Agricultural Water-Saving Strategy: Analysis of Water Use Efficiency Based on an SE-DEA Model Conducted in Xi'an, Shaanxi Province. *Water Sci. Technol.* 74, 1106–1115. doi:10.2166/wst.2016.286

Nataraja, N. R., and Johnson, A. L. (2011). Guidelines for Using Variable Selection Techniques in Data Envelopment Analysis. *Eur. J. Oper. Res.* 215 (3), 662–669. doi:10.1016/j.ejor.2011.06.045

Othman, F. M., Mohd-Zamil, N. A., Rasid, S. Z. A., Vakilbashi, A., and Mokhber, M. (2016). Data Envelopment Analysis: A Tool of Measuring Efficiency in Banking Sector. *Int. J. Econ. Financ. Issues* 6, 911–916.

Qi, Q., and Song, S. (2020). Measurement and Influencing Factors of Industrial Water Resource Utilization Efficiency in Yangtze River Economic Belt. *Int. J. Des. Nat. Ecodynamics* 15 (5), 653–658. doi:10.18280/ijdne.150506

Qiu, Y. Y., and Sheng, J. (2020). Study on Urban Water Resources Utilization Efficiency Based on SE-DEA and Malmquist Index. *J. North China Univ. Water Resour. Electric Power (Natural Sci. Edition)* 41 (05), 25–33. doi:10.19760/j.ncwu.zk.2020059 (in Chinese).

Sala-Garrido, R., Molinos-Senante, M., and Hernández-Sancho, F. (2012). How Does Seasonality Affect Water Reuse Possibilities? an Efficiency and Cost Analysis. *Resour. Conservation Recycling* 58, 125–131. doi:10.1016/j.resconrec.2011.11.002

Salehi, M. (2022). Global Water Shortage and Potable Water Safety; Today's Concern and Tomorrow's Crisis. *Environ. Int.* 158, 106936. doi:10.1016/j.envint.2021.106936

Shi, C., Zeng, X., Yu, Q., Shen, J., and Li, A. (2021). Dynamic Evaluation and Spatiotemporal Evolution of China's Industrial Water Use Efficiency Considering Undesirable Output. *Environ. Sci. Pollut. Res.* 28, 20839–20853. doi:10.1007/s11356-020-11939-x

Shi, Z., Huang, H., Wu, F., Chiu, Y.-h., and Zhang, C. (2020). The Driving Effect of Spatial Differences of Water Intensity in China. *Nat. Resour. Res.* 29, 2397–2410. doi:10.1007/s11053-019-09602-5

Siampour, L., Vahdatpour, S., Jahangiri, M., Mostafaeipour, A., Goli, A., Shamsabadi, A. A., et al. (2021). Techno-Enviro Assessment and Ranking of Turkey for Use of Home-Scale Solar Water Heaters. *Sustainable Energ. Tech. Assessments* 43, 100948. doi:10.1016/j.seta.2020.100948

SLT (2020a). Department of Water Resources of Guangdong Province- Overview of Water Resources. Available at: (in Chinese)http://slt.gd.gov.cn/szy8924/content/post_888844.html (Accessed 5 8, 2020).

SLT (2020b). Department of Water Resources of Guangdong Province- Overview of Water Resources. Available at: (in Chinese)http://slt.gd.gov.cn/szy8924/content/post_888844.html (Accessed 04 30, 2020).

SOHU (2020). Water Shortage in Guangdong-Hong Kong-Macao Bay Area-Water Resources. Available at: (in Chinese)www.sohu.com/a/404225825_120705352 (Accessed May 8, 2021).

Sun, B., and Yang, X. (2019). Analysis of Water Resources Utilization Efficiency in Jilin Province Based on DEA Method. *Destech Trans. Comput. Sci. Eng.*, 700–704. doi:10.12783/dtcse/iteee2019/28828

Sun, C. Z., and Liu, Y. Y. (2009). Analysis of the Spatial-Temporal Pattern of Water Resources Utilization Relative Efficiency Based on DEA-ESDA in China. *Resour. Sci.* 31 (10), 1696–1703. doi:10.3321/j.issn:1007-7588.2009.10.010 (in Chinese).

Sun, Y., Ren, F., Liu, J., Shi, N., Wang, H., and You, X. (2021). Evaluation of Wastewater Pollution and Treatment Efficiencies in China during Urbanization Based on Dynamic Exogenous Variable Data Envelopment Analysis. *Front. Environ. Sci.* 9, 585718. doi:10.3389/fenvs.2021.585718

Wan, Y. H., and Zhang, G. Y. (2012). Thoughts on Accelerating the Development and Utilization of Unconventional Water Sources. *China Water Resour.* 17, 9–10. doi:10.3969/j.issn.1000-1123.2012.17.004 (in Chinese).

Wang, M., Huang, Y., and Li, D. (2021). Assessing the Performance of Industrial Water Resource Utilization Systems in China Based on a Two-Stage DEA Approach with Game Cross Efficiency. *J. Clean. Prod.* 312, 127722. doi:10.1016/j.jclepro.2021.127722

Wang, Q., Jiang, R., and Li, R. (2018). Decoupling Analysis of Economic Growth from Water Use in City: A Case Study of Beijing, Shanghai, and Guangzhou of China. *Sustain. Cities Soc.* 41, 86–94. doi:10.1016/j.scs.2018.05.010

Wang, S., Zhou, L., Wang, H., and Li, X. (2018). Water Use Efficiency and its Influencing Factors in China: Based on the Data Envelopment Analysis (DEA)-Tobit Model. *Water* 10 (7), 8–32. doi:10.3390/w10070832

Wu, H. A., Zeng, B., and Zhou, M. (2017). Forecasting the Water Demand in Chongqing, China Using a Grey Prediction Model and Recommendations for the Sustainable Development of Urban Water Consumption. *Int. J. Environ. Res. Public Health* 14 (11), 1386. doi:10.3390/ijerph14111386

Zhang, G. J., and Wu, H. Q. (2020). Measurement and the Spatial Interaction Analysis of the Water Resource Comprehensive Utilization Efficiency in China. *J. Quantitative Tech. Econ.* 37 (08), 123–139. doi:10.13653/j.cnki.jqte.2020.08.007 (in Chinese).

Zhu, D., and Tang, L. (2020). Efficiency Assessment of Water Resource Utilization in the Chinese Provincial Capital Cities Based Om Data Envelopment analysis. *Acta Ecologica Sinica* 40 (06), 1–11. doi:10.5846/stxb201903020385 (in Chinese).

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Earth Science

ORIGINAL RESEARCH
published: 10 March 2022
doi: 10.3389/feart.2022.816055

Quantitative Evaluation and Diagnosis of Water Resources Carrying Capacity (WRCC) Based on Dynamic Difference Degree Coefficient in the Yellow River Irrigation District

Yi Cui[1,2], *Yuliang Zhou*[1,2], *Juliang Jin*[1,2]*, *Chengguo Wu*[1,2], *Libing Zhang*[1,2] and *Shaowei Ning*[1,2]

[1]School of Civil Engineering, Hefei University of Technology, Hefei, China, [2]Institute of Water Resources and Environmental Systems Engineering, Hefei University of Technology, Hefei, China

In order to effectively deal with the uncertainty between evaluation samples and evaluation criteria, and quantitatively identify the water resources carrying capacity (WRCC) and its obstacle factors in the Yellow River irrigation district, a calculation method of dynamic difference degree coefficient varying with evaluation sample was proposed, and then an evaluation and diagnosis model of WRCC was established. The results applied to the Dagong irrigation district showed that the overall WRCC of five counties in the irrigation district were improved from 2010 to 2017, especially since 2013. The improvement magnitudes of Changyuan County, Fengqiu County, and Hua County were significantly higher than those of Xun County and Neihuang County. In 2017, Fengqiu County, Changyuan County, Hua County, Xun County, and Neihuang County were in water resources critical overloaded status, and the connection number values were 0.231, 0.163, 0.120, –0.293, and –0.331, respectively, which is consistent with the fact that their distances become farther from the main stream of the Yellow River. In addition, the utilization ratio of water resources, available water resources amount per capita, GDP per capita, and water deficient ratio in each county belonged to the middle or strong obstacle index over a long period of time. They were the crucial obstacle factors of WRCC in the Dagong irrigation district, as well as the core and difficult points of water resources management. In some counties, the effective irrigation area ratio, effective utilization coefficient of irrigation water, and water consumption ratio of the ecological environment gradually developed from strong obstacle to weak or strong promotion index. These were important reasons for the improvement of their carrying situation, reflecting their control of agricultural and ecological water consumption. In short, the results of the case study suggest that the model established in this study is conducive to the identification of water resources' carrying status and its key obstacle factors in the Yellow River irrigation district, and can be applied to the evaluation and regulation of resources and environment carrying capacity.

Keywords: water resources carrying capacity evaluation, obstacle factor diagnosis, Yellow River irrigation district, dynamic difference degree coefficient, connection number, set pair analysis

OPEN ACCESS

Edited by:
Fei Tian,
China Agricultural University, China

Reviewed by:
Weili Duan,
Xinjiang Institute of Ecology and Geography (CAS), China
Dengfeng Liu,
Xi'an University of Technology, China

***Correspondence:**
Juliang Jin
JINJL66@126.com

Specialty section:
This article was submitted to "Environmental Informatics and Remote Sensing", a section of the journal Frontiers in Earth Science

Received: *16 November 2021*
Accepted: *31 January 2022*
Published: *10 March 2022*

Citation:
Cui Y, Zhou Y, Jin J, Wu C, Zhang L and Ning S (2022) Quantitative Evaluation and Diagnosis of Water Resources Carrying Capacity (WRCC) Based on Dynamic Difference Degree Coefficient in the Yellow River Irrigation District. Front. Earth Sci. 10:816055. doi: 10.3389/feart.2022.816055

INTRODUCTION

The Yellow River irrigation district plays a key role in national food security and regional water security in China (Ren et al., 2018; Xiong et al., 2021; Yin et al., 2021). However, due to the conflicting issues in water resources supply and demand, in combination with a fragile ecological environment and severe water problems in the Yellow River basin, the sustainable development of the Yellow River irrigation district is significantly restricted at present (Gonçalves et al., 2007; Jia et al., 2013; Miao et al., 2015). Water resources carrying capacity (WRCC) is a key index to quantify the sustainability of regional water resources utilization. Therefore, accurate evaluation and diagnosis of obstacle factors for WRCC in the Yellow River irrigation district are of great significance to ensure the coordinated development of water resources, the social economy, and the ecological environment (Pereira et al., 2007; Zhang X. et al., 2020; Jin et al., 2021).

Over the past 3 decades, WRCC has become a hot and difficult issue in the field of sustainable water resources utilization (Wang et al., 2013; Peng et al., 2021; Qi et al., 2021). Gong and Jin (2009) established a fuzzy comprehensive evaluation model of WRCC. Wang et al. (2018) constructed an evaluation model of WRCC based on entropy and synergy theories. Similarly, Dai et al. (2019) used a system dynamics model to evaluate WRCC. Song et al. (2020) assessed WRCC using the catastrophe series method. Wu et al. (2020) built an evaluation model of WRCC based on cloud model. Furthermore, Zhao et al. (2021) established an evaluation and influencing factor analysis model of WRCC based on the theoretical framework of pressure support, damage recovery, and recession promotion. However, previous studies on WRCC evaluation have mostly focused on cities or regions, while those focusing on irrigation districts are scarce (Kang et al., 2019; Zhang et al., 2019; He et al., 2021). In addition, research on the obstacle factor diagnosis of WRCC is even less common (Cui et al., 2018). Therefore, it is urgent to establish an effective evaluation and diagnosis model, as well as accurately identify the level and obstacle factors of WRCC in the Yellow River irrigation district.

The WRCC system is affected by many factors, including water resources, the social economy, and the ecological environment, and is a typically complex system (Yang et al., 2015; Wang et al., 2019; Liao et al., 2020). At present, the multi-index system comprehensive evaluation is an effective method of WRCC evaluation by constructing an evaluation index system and evaluation model (Zhang et al., 2019; Peng and Deng, 2020; Zuo et al., 2021). However, most models are unable to fully consider the uncertainty between evaluation samples and evaluation criteria, resulting in the deviations of evaluation results (Jin et al., 2008). Set pair analysis (SPA) is a new system uncertainty analysis method which fully reflects the certainty and uncertainty relationships between evaluation samples and evaluation criteria by the same-different-inverse structure of connection number for the set pair constructed by these two sets (Zhao, 2000; Kumar and Garg, 2016; Chong et al., 2017). This method has been widely used to comprehensively evaluate water resources system problems (Wang et al., 2009; Li et al., 2019; Zhang et al., 2021). Yang et al. (2014) used SPA to evaluate the vulnerability of water resources system. Roy and Datta (2019) studied the adaptive management of coastal aquifers based on SPA and entropy theory. In addition, Su et al. (2019) conducted water security evaluation using SPA and scenario simulation. Similarly, Lyu et al. (2021) coupled SPA and fuzzy number to assess the risk of urban water quality. Wan et al. (2021) built a river health evaluation and prediction model based on SPA and extension theory. Nevertheless, the majority of studies only reflect the evaluation results based on the connection number components (Li et al., 2021). Thus, it is difficult to obtain the value of the complete connection number, which results in a loss of information, thereby limiting the development of SPA. Moreover, the key is to reasonably determine the difference degree coefficient of the connection number.

As a key part of the connection number, the difference degree coefficient is used to quantitatively describe the uncertainty of the constructed set pair at the micro level (Zhao, 2000; Wang et al., 2009; Pan et al., 2017). This has a significant impact on the evaluation results. Several researchers have reported on the calculation methods of the difference degree coefficient. For example, Li et al. (2009) deduced and optimized the difference degree coefficient using the target value of the connection number. Tang (2009) proposed an expert estimation method of the difference degree coefficient. Furthermore, Pan et al. (2016) and Li et al. (2021) respectively constructed the trapezoidal and triangular fuzzy numbers of the difference degree coefficient, and determined the variation range of difference degree coefficient at a given cut set level. Jin et al. (2019a, 2019b) allocated the difference degree coefficient in proportions using grey correlation degree and full partial certainty, respectively. However, the majority of these methods are relatively rough and only obtain an approximate value or an interval range of difference degree coefficients (Tang, 2009; Pan et al., 2016; Li et al., 2021). In fact, accurately calculating the value of the difference degree coefficient remains difficult, which leads to large deviations between the research results and reality. In addition, this coefficient should be determined in combination with evaluation sample information. Therefore, at present, there is a need for the development of an effective method with which to calculate the difference degree coefficient.

In this study, based on the theoretical analysis and practical investigation of the water resources carrying characteristics in the Yellow River irrigation district, an evaluation index system and evaluation grade criteria were built. Additionally, a method to calculate the dynamic difference degree coefficient varying with the evaluation sample was also proposed, and the value of the complete connection number was obtained. Then, a model was established to quantitatively evaluate the WRCC and diagnose its key obstacle factors in the Yellow River irrigation district. Furthermore, this model was further applied to the Dagong irrigation district in Henan Province, China. The results provide scientific support for water resources management and decision-making in the Yellow River irrigation district.

MATERIALS AND METHODS

Evaluation and Diagnosis Model of WRCC Based on Dynamic Difference Degree Coefficient

The evaluation and diagnosis model of WRCC in the Yellow River irrigation district based on dynamic difference degree coefficient was constructed according to the following steps (**Figure 1**):

***Step* 1:** Based on the structural and functional analysis of WRCC system (Jin et al., 2018; Liao et al., 2020), combined with the results of practical study, expert consultation, and literature review in the Yellow River irrigation district (Zhang et al., 2019; Zhang X. Y. et al., 2020; Jin et al., 2021), an evaluation index system $\{x_{kj}|k = 1, 2, 3; j = 1, 2, \ldots, n_k\}$ was established. The evaluation index sample set was set as $\{x_{ikj}|i = 1, 2, \ldots, m; k = 1, 2, 3; j = 1, 2, \ldots, n_k\}$, where x_{ikj} denotes the value of index j in subsystem k for sample i, m denotes the number of evaluation samples, $k = 1, 2, 3$ denotes the water resources carrying support force subsystem, carrying pressure force subsystem, and carrying regulation force subsystem, respectively, and n_k denotes the number of indexes in subsystem k. In addition, in this study, the WRCC in the Yellow River irrigation district was divided into three evaluation grades (Cui et al., 2018; Li et al., 2021) $\{s_{gkj}|g = 1, 2, 3; k = 1, 2, 3; j = 1, 2, \ldots, n_k\}$, where $g = 1, 2, 3$ denotes the water resources loadable status, critical overloaded status, and overloaded status, respectively.

***Step* 2:** The improved fuzzy analytical hierarchy process based on accelerating genetic algorithm (AGA-FAHP) (Jin et al., 2004) was used to determine the weight of subsystem and that of each evaluation index $\{w_{kj}|k = 1, 2, 3; j = 1, 2, \ldots, n_k\}$.

For subsystem k, experts were invited to compare the importance of each index in this subsystem to the WRCC in the irrigation district, and the fuzzy complementary judgment matrix $A^k = (a^k_{jl})_{nk\times nk}$ was obtained. This matrix met $0 \le a^k_{jl} \le 1$ and $a^k_{jl} + a^k_{lj} = 1$, $a^k_{jl} = 0.5$, indicating that index j was as important as index l, $a^k_{jl} > 0.5$, indicating that index j was more important than index l, and vice versa. Furthermore, the AGA-FAHP was

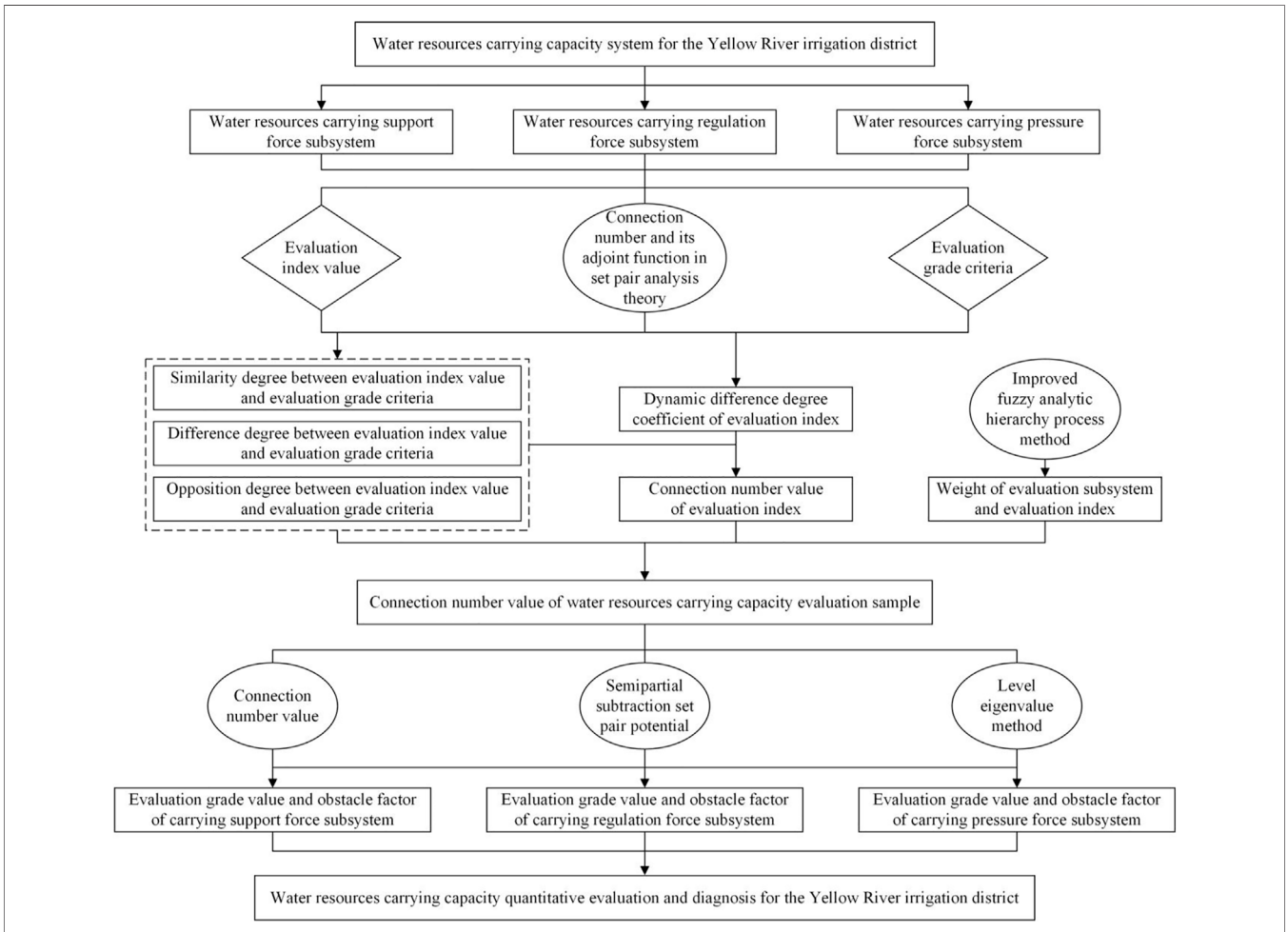


FIGURE 1 | Construction process of evaluation and diagnosis model for water resources carrying capacity (WRCC) in the Yellow River irrigation district based on dynamic difference degree coefficient.

used to verify and modify the consistency of A^k and calculate the weight of each evaluation index w_{kj}. If A^k satisfies additive transitivity, then (Song and Yang, 2003):

$$\left(a_{jq}^k - 0.5\right) + \left(a_{ql}^k - 0.5\right) = \left(a_{jl}^k - 0.5\right), \quad k = 1, 2, 3;\ j, q, l = 1, 2, \ldots, n_k \tag{1}$$

where A^k denotes the fuzzy consistency judgment matrix, (a_{jl}^k – 0.5) describes the importance of index j compared with index l, and the fuzzy consistency judgment matrix meant that this preference of importance could be transmitted. Moreover, if A^k meets complete consistency, then (Song and Yang, 2003):

$$\sum_{j=1}^{n_k}\sum_{l=1}^{n_k}\left|0.5\left(n_k - 1\right)\left(w_{kj} - w_{kl}\right) + 0.5 - a_{jl}^k\right| \Big/ n_k^2 = 0 \tag{2}$$

where the left term in **Eq. 2** is the consistency index of A^k. If this index did not exceed a certain critical value, it would indicate that A^k had satisfactory consistency. Otherwise, A^k should be modified. The modified matrix was $B^k = (b_{jl}^k)_{nk\times nk}$, and the sorting weight of the element in B^k was still set as $\{w_{kj}|k = 1, 2, 3;\ j = 1, 2, \ldots, n_k\}$. Hence, B^k satisfies the following equation (Jin et al., 2004):

$$\begin{aligned}\min CIC\left(n_k\right) = &\sum_{j=1}^{n_k}\sum_{l=1}^{n_k}\left|b_{jl}^k - a_{jl}^k\right| \Big/ n_k^2 + \sum_{j=1}^{n_k}\sum_{l=1}^{n_k}\left|0.5\left(n_k - 1\right)\left(w_{kj} - w_{kl}\right)\right.\\ &\left. + 0.5 - b_{jl}^k\right| \Big/ n_k^2\end{aligned}$$

$$\text{s.t.}\quad \begin{cases} b_{jj}^k = 0.5, \quad k = 1, 2, 3;\ j = 1, 2, \ldots, n_k \\ 1 - b_{lj}^k = b_{jl}^k \in \left[a_{jl}^k - d, a_{jl}^k + d\right] \cap [0, 1], \\ \qquad k = 1, 2, 3;\ j = 1, 2, \ldots, n_k;\ l = j + 1, j + 2, \ldots, n_k \\ \sum_{j=1}^{n_k} w_{kj} = 1.0, w_{kj} \in [0, 1], \quad k = 1, 2, 3;\ j = 1, 2, \ldots, n_k \end{cases} \tag{3}$$

which represents the optimal fuzzy consistency judgment matrix for A^k. Here, $CIC(n_k)$ is the consistency index coefficient and d is a parameter that usually takes the value within [0, 0.5]. The sorting weight $\{w_{kj}|k = 1, 2, 3;\ j = 1, 2, \ldots, n_k\}$ and the upper triangular elements in B^k were the optimization variables. Accelerating genetic algorithm (AGA) is a general global optimization method and it is effective in solving **Eq. 3**. In addition, when $CIC(n_k)$ was less than a certain critical value, A^k was considered to have satisfactory consistency, and the obtained sorting weight of each evaluation index was acceptable. Otherwise, d should be adjusted. Based on a large number of simulation tests and relevant studies (Jin et al., 2004; Cui et al., 2018), a critical value of 0.20 was selected in this study.

***Step* 3:** SPA is an original theory of uncertainty quantitative analysis proposed by Chinese scholar Zhao Keqin in 1989 (Zhao, 2000). The foundation of SPA was the connection number u of the set pair constructed by two sets with a common attribute, and the certainty and uncertainty relationships of the set pair were quantitatively described by a same-different-inverse structure. Specifically, according to SPA, the certainty of the set pair was divided into two aspects of similarity and opposition, which were quantitatively expressed by a and c, respectively. Furthermore, the uncertainty of the set pair was divided into the uncertainty at a macro level, which was measured by b, and that at the micro level, which was depicted by I. The ternary connection number u, which was commonly used, is expressed as follows (Yang et al., 2014):

$$u = a + bI + cJ \tag{4}$$

where a, b, and c denote the similarity degree, difference degree, and opposition degree components of connection number u, respectively, which were used to quantitatively express the degrees of the same property, different property, and inverse property for the constructed set pair, $a + b + c = 1$. I denotes the difference degree coefficient and its value varies with the relationship type of the set pair. The set pair constructed in this study belonged to the positive-negative type, and I generally took the value within [−1, 1]. J denotes the opposition degree coefficient, which generally takes −1 for the set pair of positive-negative type (Wang et al., 2009; Cui et al., 2018).

Based on SPA and the practical problem of WRCC evaluation in the Yellow River irrigation district, two sets of evaluation index value and evaluation grade criteria constituted a set pair. According to the variable fuzzy set of the proximity between index value and grade criteria, the connection number of evaluation index was calculated. Specifically, the initial ternary connection number component u_{gikj} between WRCC evaluation index sample value x_{ikj} and grade criteria s_{gkj} (g = 1, 2, 3) in the irrigation district were, respectively, as follows (Jin et al., 2008):

$$u_{1ikj} = \begin{cases} 1, \quad s_{0kj} \le x_{ikj} \le s_{1kj} \text{ for positive index,} \\ \qquad \text{or } s_{0kj} \ge x_{ikj} \ge s_{1kj} \textbf{ for negative index} \\ 1 - 2\left(x_{ikj} - s_{1kj}\right)\Big/\left(s_{2kj} - s_{1kj}\right), \quad s_{1kj} < x_{ikj} \le s_{2kj} \textbf{ for positive index,} \\ \qquad \textbf{or } s_{1kj} > x_{ikj} \ge s_{2kj} \textbf{ for negative index} \\ -1, \quad s_{2kj} < x_{ikj} \le s_{3kj} \text{ for positive index,} \\ \qquad \text{or } s_{2kj} > x_{ikj} \ge s_{3kj} \textbf{ for negative index} \end{cases} \tag{5}$$

$$u_{2ikj} = \begin{cases} 1 - 2\left(s_{1kj} - x_{ikj}\right)\Big/\left(s_{1kj} - s_{0kj}\right), \quad s_{0kj} \le x_{ikj} \le s_{1kj} \textbf{ for positive index,} \\ \qquad \textbf{or } s_{0kj} \ge x_{ikj} \ge s_{1kj} \textbf{ for negative index} \\ 1, \quad s_{1kj} < x_{ikj} \le s_{2kj} \text{ for positive index,} \\ \qquad \text{or } s_{1kj} > x_{ikj} \ge s_{2kj} \textbf{ for negative index} \\ 1 - 2\left(x_{ikj} - s_{2kj}\right)\Big/\left(s_{3kj} - s_{2kj}\right), \quad s_{2kj} < x_{ikj} \le s_{3kj} \textbf{ for positive index,} \\ \qquad \textbf{or } s_{2kj} > x_{ikj} \ge s_{3kj} \textbf{ for negative index} \end{cases} \tag{6}$$

$$u_{3ikj} = \begin{cases} -1, \quad s_{0kj} \le x_{ikj} \le s_{1kj} \textbf{ for positive index,} \\ \qquad \textbf{or } s_{0kj} \ge x_{ikj} \ge s_{1kj} \textbf{ for negative index} \\ 1 - 2\left(s_{2kj} - x_{ikj}\right)\Big/\left(s_{2kj} - s_{1kj}\right), \quad s_{1kj} < x_{ikj} \le s_{2kj} \textbf{ for positive index,} \\ \qquad \textbf{or } s_{1kj} > x_{ikj} \ge s_{2kj} \text{ for negative index} \\ 1, \quad s_{2kj} < x_{ikj} \le s_{3kj} \textbf{ for positive index,} \\ \qquad \textbf{or } s_{2kj} > x_{ikj} \ge s_{3kj} \textbf{ for negative index} \end{cases} \tag{7}$$

where the larger the value of positive (negative) index, the higher (lower) the evaluation grade. s_{1kj} and s_{2kj} denote the critical value between grade 1 and grade 2, and that between grade 2 and grade 3, respectively, for index j in subsystem k. s_{0kj} and s_{3kj} denote the other critical values of grade 1 and grade 3, respectively.

The initial ternary connection number component u_{gikj} depended on whether the evaluation index sample value x_{ikj} belonged to the same interval, adjacent interval, or separated interval of evaluation grade g, and took 1, the value within [−1, 1], or −1, respectively. Therefore, u_{gikj} can be regarded as a relative difference degree function for the variable fuzzy set of the proximity between index value and grade criteria, the corresponding relative membership degree is as follows (Chen, 2005):

$$a^*_{ikj} = 0.5 + 0.5u_{1ikj}, b^*_{ikj} = 0.5 + 0.5u_{2ikj}, c^*_{ikj} = 0.5 + 0.5u_{3ikj}, i = 1, 2, \ldots, m; k = 1, 2, 3; j = 1, 2, \ldots, n_k \quad (8)$$

Furthermore, the final ternary connection number components a, b, and c of the WRCC evaluation index in the irrigation district were obtained using **Eq. 8** after normalization as follows (Jin et al., 2008; Cui et al., 2018):

$$a_{ikj} = a^*_{ikj} \Big/ \left(a^*_{ikj} + b^*_{ikj} + c^*_{ikj}\right),\ b_{ikj} = b^*_{ikj} \Big/ \left(a^*_{ikj} + b^*_{ikj} + c^*_{ikj}\right),\ c_{ikj} = c^*_{ikj} \Big/ \left(a^*_{ikj} + b^*_{ikj} + c^*_{ikj}\right) \quad (9)$$

***Step* 4:** According to SPA and the expression of connection number u in **Eq. 4**, the difference degree coefficient I was the link that transformed difference degree b to similarity degree a or opposition degree c under certain conditions. It was considered to be an important connector between the theoretical model of connection number and the practical research question. Therefore, how to scientifically determine I was a key and challenging problem when calculating the value of the complete connection number. However, the value process of difference degree coefficient I was highly uncertain and complex and few studies had been published at present.

The difference degree coefficient I was essential to quantitatively describe the uncertainty of set pair at the micro level (Zhao, 2000; Wang et al., 2009; Pan et al., 2017), while an important source of the uncertainty was the information carried by sample data. In addition, the physical meaning of I can be interpreted as the level of difference degree b transformed to similarity degree a or opposition degree c, wherein the direction (transformed to a or c) and magnitude should be closely related to the proximity between evaluation sample value and each evaluation grade. That was, the larger the value of a (or c), the more b transformed to a (or c). Therefore, I should vary continuously and dynamically with the practical sample value of research question. In this study, the set pair composed of WRCC evaluation index sample value x_{ikj} and grade criteria s_{gkj} in the Yellow River irrigation district belonged to the positive-negative type, where the value range of I was [−1, 1] (Zhao, 2000; Zhang, 2020a; Jin et al., 2021). Furthermore, when I was within [0, 1], this indicated that b transformed to a, whereas when I was within [−1, 0], indicating that b transformed to c, and the absolute value of I reflected the transformed magnitude. In other words, the difference degree coefficient I_{ikj} in this study continuously and dynamically changed with the evaluation index sample value x_{ikj}, as shown in **Figure 2**. Specifically, when x_{ikj} was closer to the critical value s_{0kj} of grade 1, I_{ikj} was closer to 1. When x_{ikj} was closer to the critical value s_{1kj} between grades 1 and 2, I_{ikj} was closer to 1/3. When x_{ikj} was closer to the critical value s_{2kj} between grades 2 and 3, I_{ikj} was closer to −1/3. Moreover, when x_{ikj} was closer to the critical value s_{3kj} of grade 3, I_{ikj} was closer to −1. Based on SPA and the above analysis, this study proposed a method for calculating the dynamic difference degree coefficient of the ternary connection number as follows:

$$I_{ikj} = \begin{cases} 1 - 2\left(x_{ikj} - s_{0kj}\right)\Big/3\left(s_{1kj} - s_{0kj}\right), & s_{0kj} \le x_{ij} \le s_{1kj} \textbf{ for positive index}, \\ & \textbf{or } s_{0kj} \ge x_{ikj} \ge s_{1kj} \textbf{ for negative index} \\ -2\left[x_{ikj} - \left(s_{1kj} + s_{2kj}\right)\Big/2\right]\Big/3\left(s_{2kj} - s_{1kj}\right), & s_{1kj} < x_{ikj} \le s_{2kj} \textbf{ for positive index}, \\ & \textbf{or } s_{1kj} > x_{ikj} \ge s_{2kj} \textbf{ for negative index} \\ -1 + 2\left(s_{3kj} - x_{ikj}\right)\Big/3\left(s_{3kj} - s_{2kj}\right), & s_{2kj} < x_{ikj} \le s_{3kj} \textbf{ for positive index}, \\ & \textbf{or } s_{2kj} > x_{ikj} \ge s_{3kj} \textbf{ for negative index} \end{cases} \quad (10)$$

***Step* 5:** The ternary connection number u_{ikj} of index j in subsystem k for WRCC evaluation sample i was calculated by substituting the results obtained by **Eqs. 5–10**) into **Eq. 4** (Zhao, 2000; Jia et al., 2013):

$$u_{ikj} = a_{ikj} + b_{ikj}I_{ikj} + c_{ikj}J, \quad i = 1, 2, \ldots, m; k = 1, 2, 3; j = 1, 2, \ldots, n_k \quad (11)$$

Combined with the weight w_{kj} of index j in subsystem k obtained by **Eq. 3**, the ternary connection number u_{ik} of subsystem k for sample i was calculated as follows (Jin et al., 2008):

$$u_{ik} = \sum_{j=1}^{n_k} w_{kj}a_{ikj} + \sum_{j=1}^{n_k} w_{kj}b_{ikj}I_{ikj} + \sum_{j=1}^{n_k} w_{kj}c_{ikj}J, \quad i = 1, 2, \ldots, m; k = 1, 2, 3 \quad (12)$$

Finally, the ternary connection number u_i of WRCC evaluation sample i in the Yellow River irrigation district was obtained according to the following equation (Cui et al., 2018):

$$u_i = \sum_{k=1}^{3}\sum_{j=1}^{n_k} w_k w_{kj} a_{ikj} + \sum_{k=1}^{3}\sum_{j=1}^{n_k} w_k w_{kj} b_{ikj} I_{ikj} + \sum_{k=1}^{3}\sum_{j=1}^{n_k} w_k w_{kj} c_{ikj} J, \quad i = 1, 2, \ldots, m \quad (13)$$

where w_k denotes the weight of subsystem k, which can be calculated using the AGA-FAHP.

The evaluation grade value of WRCC in the irrigation district was calculated based on the connection number value u obtained by **Eq. 13**, wherein $u \in [-1, 1]$. Furthermore, u was divided into three levels according to the critical values of evaluation index grade criteria, which corresponded to water resources overloaded status $u \in [-1.000, -0.667)$, critical overloaded status $u \in [-0.667, 0.667]$, and loadable status $u \in (0.667, 1.000]$, respectively.

To compare with the connection number value u, the evaluation grade value h_i of sample i was calculated using the level eigenvalue method (Zhou et al., 2022) in **Eq. 14**. h was also divided into three levels according to the critical values of evaluation index grade criteria, which corresponded to water resources overloaded status $h \in (2.5, 3.0]$, critical overloaded status $h \in [1.5, 2.5]$, and loadable status $h \in [1.0, 1.5)$, respectively.

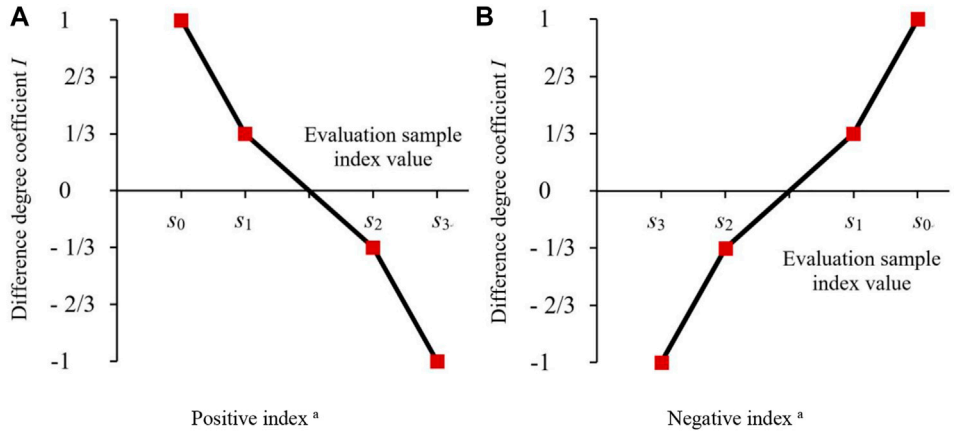


FIGURE 2 | Value process for dynamic difference degree coefficient of the ternary connection number changes with evaluation index sample value. The larger the value of positive (negative) index, the higher (lower) the evaluation grade.

$$h_i = \sum_{k=1}^{3}\sum_{j=1}^{n_k} w_k w_{kj} a_{ikj} + 2\sum_{k=1}^{3}\sum_{j=1}^{n_k} w_k w_{kj} b_{ikj} + 3\sum_{k=1}^{3}\sum_{j=1}^{n_k} w_k w_{kj} c_{ikj}, \quad i = 1, 2, \ldots, m \tag{14}$$

***Step 6*:** The main obstacle factors hindering the improvement of WRCC in the Yellow River irrigation district were diagnosed based on the connection number value u_{ikj} of evaluation index calculated using **Eq. 11**. It can be proven that $u_{ikj}\in[-1, 1]$, and based on the principle of equal division, the evaluation indexes were divided into five types. They were strong obstacle $u_{ikj}\in[-1.0, -0.6)$, middle obstacle $u_{ikj}\in[-0.6, -0.2)$, weak obstacle $u_{ikj}\in[-0.2, 0.2]$, weak promotion $u_{ikj}\in(0.2, 0.6]$, and strong promotion indexes $u_{ikj}\in(0.6, 1.0]$. Moreover, the strong or middle obstacle index was the factor that seriously hindered the improvement of WRCC, and at the same time, the key focus of water resources regulation and control in the irrigation district.

For comparison with the connection number value u, the set pair potential of connection number for evaluation index was used to diagnose the obstacle factors of WRCC. According to SPA, set pair potential reflected the overall development trend of set pair at the macro level (Zhao, 2000; Zhou et al., 2022). To quantitatively describe this trend, Cui et al. (2018) constructed the subtraction set pair potential $s_1(u)$ of the ternary connection number in **Eq. 15**, the basic idea was to allocate b to a and c according to the proportions of $a/(a+b+c)$ and $c/(a+b+c)$, respectively:

$$s_1(u_{ikj}) = \left[a_{ikj} + b_{ikj}a_{ikj}/(a_{ikj}+b_{ikj}+c_{ikj})\right] - \left[c_{ikj} + b_{ikj}c_{ikj}/(a_{ikj}+b_{ikj}+c_{ikj})\right] \tag{15}$$

In addition, the partial connection number reflected the overall development trend of set pair at the micro level (Zhao, 2000). From the perspective of same-different-inverse transformation based on the partial connection number, to transform the similarity component of b to a, $(a+b)$ should be regarded as a whole of the similarity and should not include the opposition degree c. Therefore, it was more reasonable to take $a/(a+b)$ as the proportion of b allocated to a than $a/(a+b+c)$ in **Eq. 15**. Similarly, to transform the opposition component of b to c, $(c+b)$ should be regarded as a whole of the opposition and should not include the similarity degree a. It was more reasonable to consider $c/(c+b)$ as the proportion of b allocated to c than $c/(a+b+c)$ in **Eq. 15**. Therefore, based on the subtraction set pair potential and further combined with the idea of the partial connection number, a new adjoint function of the ternary connection number, the semipartial subtraction set pair potential $s_2(u)$ was proposed (Jin et al., 2021):

$$s_2(u_{ikj}) = \begin{cases} -1, & a_{ikj} + b_{ikj} = 0 \\ 1, & b_{ikj} + c_{ikj} = 0 \\ \left[a_{ikj} + b_{ikj}a_{ikj}/(a_{ikj}+b_{ikj})\right] - \left[c_{ikj} + b_{ikj}c_{ikj}/(c_{ikj}+b_{ikj})\right], & a_{ikj} + b_{ikj} \neq 0 \text{ and } b_{ikj} + c_{ikj} \neq 0 \end{cases} \tag{16}$$

It can be seen that $s_2(u_{ikj})\in[-1, 1]$, according to the principle of equal division, $s_2(u_{ikj})$ was divided into five levels. They were inverse potential $s_2(u_{ikj})\in[-1.0, -0.6)$, partial inverse potential $s_2(u_{ikj})\in[-0.6, -0.2)$, symmetrical potential $s_2(u_{ikj})\in[-0.2, 0.2]$, partial identical potential $s_2(u_{ikj})\in(0.2, 0.6]$, and identical potential $s_2(u_{ikj})\in(0.6, 1.0]$, respectively. Furthermore, the index of partial inverse potential or inverse potential was diagnosed as the main obstacle factor hindering the improvement of WRCC and also the important object of water resources management in the Yellow River irrigation district (Li et al., 2021).

Study Area

The evaluation and diagnosis model of WRCC based on dynamic difference degree coefficient was applied to the Dagong ecological irrigation district in Henan Province, China. The level of WRCC and key obstacle factors were quantitatively identified in the typical Yellow River irrigation district. The Dagong irrigation district was located in the north of the Yellow River (**Figure 3**), with a design irrigation area of 190,000 ha and a land area of

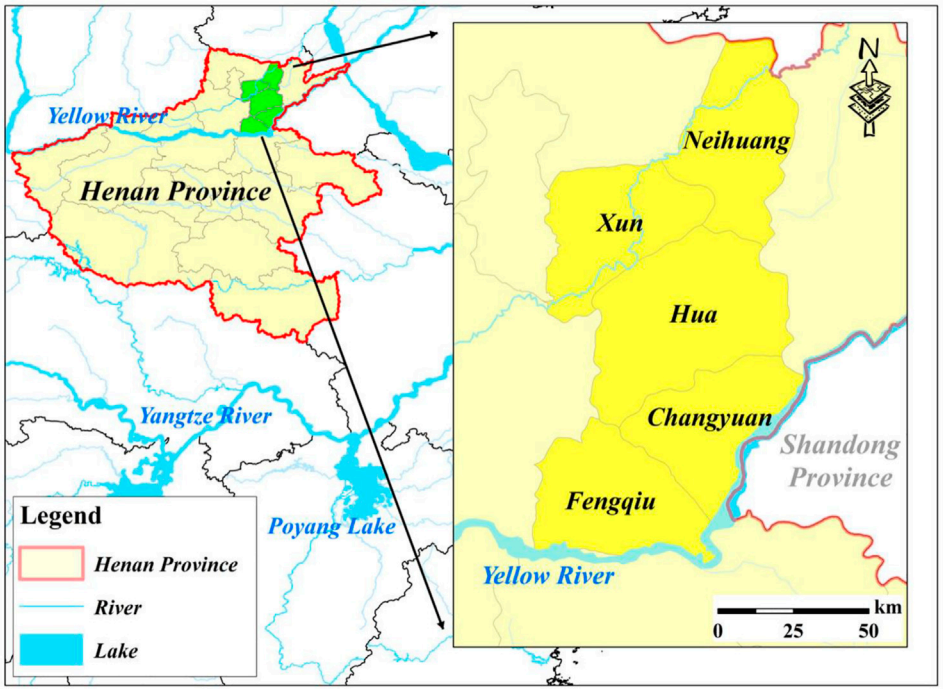


FIGURE 3 | Location of the Dagong ecological Yellow River irrigation district in Henan Province, China.

2,886 km^2. It was a large-scale Yellow River irrigation district in Henan Province (Zhang et al., 2019; Du et al., 2020). This irrigation district included Neihuang County, Xun County, Hua County, Changyuan County, and Fengqiu County, all of these were advanced counties of grain production in China. The main planting crops included wheat, rice, and corn, while intercropping crops included peanuts, millet, beans, cotton, and sesame (Zhang X. et al., 2020). In recent years, with the rapid development of the social economy in the irrigation district, the demand for water resources for agriculture, life, secondary and tertiary industries, and ecology had continuously increased, which resulted in a prominent contradiction between limited supply and increased demand for water resources (Gonçalves et al., 2007; Pereira et al., 2007; Jia et al., 2013). In summary, the water resources situation in the Dagong irrigation district was severe.

RESULTS AND DISCUSSION

Establishment of WRCC Evaluation Index System in Irrigation District

According to the evaluation target and construction principle of evaluation index system (Gong and Jin, 2009; Song et al., 2011; Chang et al., 2020), and combined with the influencing factors of the real process for WRCC in the ecological Yellow River irrigation district, the WRCC system was divided into three subsystems. They were water resources carrying support force subsystem, carrying pressure force subsystem, and carrying regulation force subsystem, respectively (Cui et al., 2018; Jin et al., 2018; Li et al., 2021). Moreover, based on the comprehensive analysis of carrying characteristics for these three subsystems, the actual development of water resources, the social economy, and the ecological environment in the Dagong irrigation district, as well as relevant research (Du et al., 2020; Zhang X. et al., 2020; Jin et al., 2021), an evaluation index system (three subsystems, a total of nine indexes X_1—X_9), and the corresponding evaluation grade criteria (water resources loadable, critical overloaded, and overloaded status) are listed in **Table 1**. In addition, the weights of subsystem and each evaluation index were determined using the AGA-FAHP in **Eq. 3**, together with relevant studies (Du et al., 2020; Zhang X. et al., 2020) (**Table 1**).

In this study, the sample data of WRCC evaluation index in the Dagong irrigation district were obtained from the Henan Water resources Bulletin (2010–2017), the Henan Statistical Yearbook (2011–2018), and phased achievements of the third national survey and evaluation of water resources utilization in Henan Province, China.

Evaluation and Analysis of WRCC for Five Counties in Irrigation District

The sample data of nine WRCC evaluation indexes for five counties in the Dagong irrigation district from 2010 to 2017, and the corresponding evaluation grade criteria in **Table 1**, were substituted into **Eq. 5–9** to obtain the connection number components of each index. Furthermore, the dynamic difference degree coefficient of the connection number for

TABLE 1 | Evaluation index system and evaluation grade criteria of water resources carrying capacity (WRCC) in the ecological Yellow River irrigation district (Du et al., 2020; Zhang X. et al., 2020; Jin et al., 2021).

Evaluation system	Evaluation subsystem	Evaluation index	Evaluation grade criteria			Index weight	Index type[a]
			Grade 1 (loadable)	Grade 2 (critical overloaded)	Grade 3 (overloaded)		
WRCC system in the ecological Yellow River irrigation district	Carrying support force	Available water resources amount per capita X_1 (m^3)	≥500	[400, 500)	<400	0.041	Negative
		Utilization ratio of water resources X_2 (%)	≤40	(40, 60]	>60	0.166	Positive
	Carrying pressure force	Average urbanization ratio X_3 (%)	≤35	(35, 40]	>40	0.028	Positive
		GDP per capita X_4 (Yuan)	≥50000	[30000, 50000)	<30000	0.083	Negative
		Effective irrigation area ratio X_5 (%)	≥60	[40, 60)	<40	0.186	Negative
		Water deficient ratio X_6 (%)	≤10	(10, 20]	>20	0.028	Positive
		Shallow groundwater exploitation ratio X_7 (%)	≤10	(10, 17.5]	>17.5	0.055	Positive
	Carrying regulation force	Effective utilization coefficient of irrigation water X_8	≥0.65	[0.60, 0.65)	<0.60	0.166	Negative
		Water consumption ratio of ecological environment X_9 (%)	≥5	[3, 5)	<3	0.247	Negative

[a]The larger the value of positive (negative) index, the higher (lower) the evaluation grade.

each index was calculated in combination with **Figure 2** and **Eq. 10**. Then, according to the index weight in **Table 1** and **Eq. 11–13**, the connection number values of evaluation index, subsystem, and WRCC were calculated. Meanwhile, the evaluation grade value according to the level eigenvalue method was obtained using **Eq. 14**. Finally, the evaluation results of WRCC in the Dagong irrigation district are shown in **Figures 4–6**.

(1) According to the average value of WRCC evaluation connection number u from 2010 to 2017 for each county (**Figure 4**), from the perspective of multi-year average, Xun County and Neihuang County belonged to the water resources overloaded status ($-1.000 \le u < -0.667$), while the other three counties were in critical overloaded status ($-0.667 \le u \le 0.667$). The results are consistent with those of Du et al. (2020) and Zhang et al. (2021). It indicates that although the carrying status of the five counties has been different over the years, the overall carrying situation in the Dagong irrigation district is serious. Thus, there is a need to analyze the driving mechanism of WRCC in different counties, identify the key factors hindering the improvement of WRCC, and then take the corresponding measures.

In addition, it can be seen that from the average value of the connection number component (**Figure 4**), the multi-year average values of opposition degree c for the five counties were all significantly larger than those of similarity degree a, which was consistent with the above evaluation result of severe water resources carrying situation for these five counties. Moreover, the average values of a in Hua County, Fengqiu

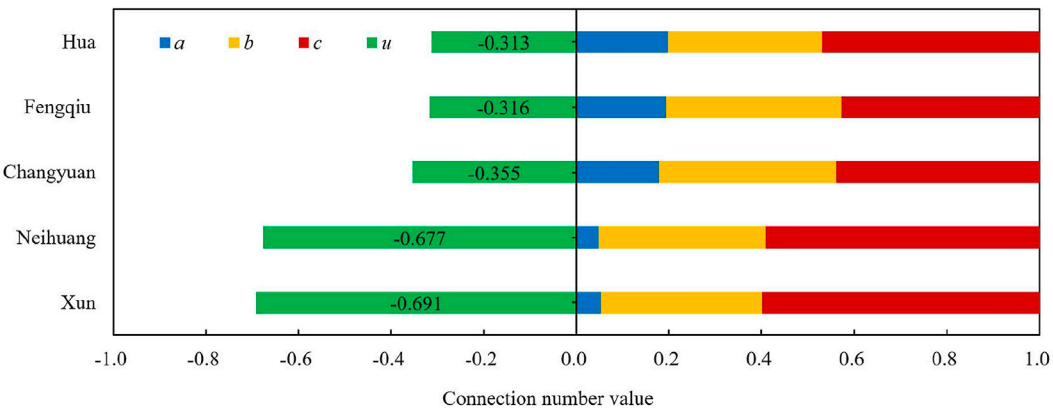


FIGURE 4 | Average values of connection number u and its components a, b, and c for water resources carrying capacity (WRCC) evaluation from 2010 to 2017 in the Dagong irrigation district.

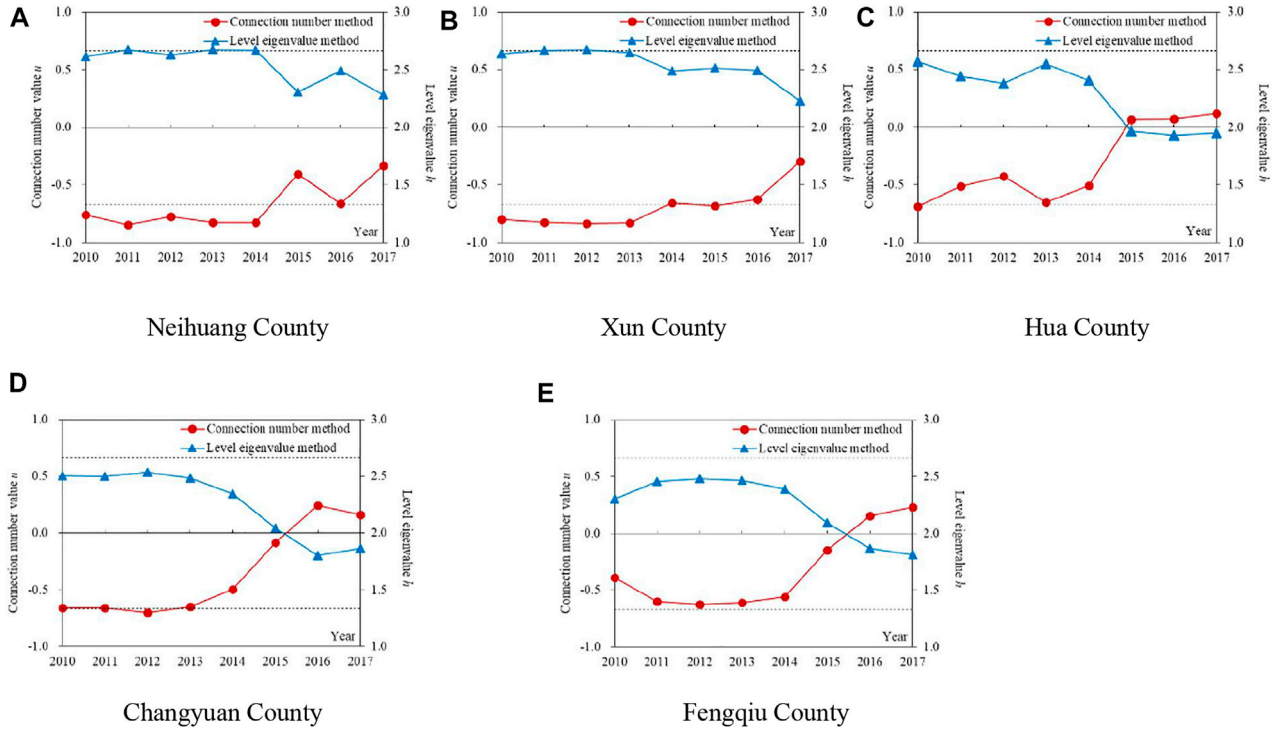


FIGURE 5 | Evaluation grade values of two methods for water resources carrying capacity (WRCC) from 2010 to 2017 in the Dagong irrigation district. The two dashed lines from top to bottom in the figures represent the boundary of connection number value for water resources loadable and critical overloaded status (0.667), and that for critical overloaded and overloaded status (–0.667) in each county, respectively.

County, and Changyuan County were relatively higher (0.198, 0.193, and 0.179, respectively), while those of *c* in Xun County and Neihuang County were relatively larger (0.598 and 0.591) (**Figure 4**). These results were consistent with the above result that the WRCC in Hua County, Fengqiu County, and Changyuan County were stronger than those in Xun County and Neihuang County.

(2) It can be seen from the evaluation results of two methods (**Figure 5**), the changing trends of the evaluation grade values of WRCC for five counties from 2010 to 2017 calculated by connection number method were consistent with those of level eigenvalue method. Furthermore, the changing range and the differences in grade value year to year obtained by connection number were more significant than level eigenvalue. It shows that the evaluation method of the connection number based on the dynamic difference degree coefficient can fully excavate and make use of the information carried by sample data. Furthermore, the uncertainty of the WRCC system can be accurately quantified by the dynamic difference degree coefficient varying with the actual evaluation sample, and the value of the complete connection number can be directly determined. Therefore, the WRCC evaluation method proposed in this study has a higher sensitivity and accuracy, and the evaluation results are reasonable and reliable.

(3) According to the connection number values and the corresponding grade values of WRCC evaluation for each county from 2010 to 2017 (**Figures 6**, **7**), from the perspective of overall development in the irrigation district, from 2010 to 2017, most counties were in water resources overloaded status ($-1.000 \leq u < -0.667$), some belonged to critical overloaded status ($-0.667 \leq u \leq 0.667$), and others were in loadable status ($0.667 < u \leq 1.000$). Furthermore, the number of overloaded counties decreased from three in 2010 to zero in 2017, and the average value of the connection number for the five counties decreased from –0.656 in 2010 to –0.022 in 2017. After 2013, the average value from 2010 to 2013 and that from 2014 to 2017 were –0.681 (overloaded) and –

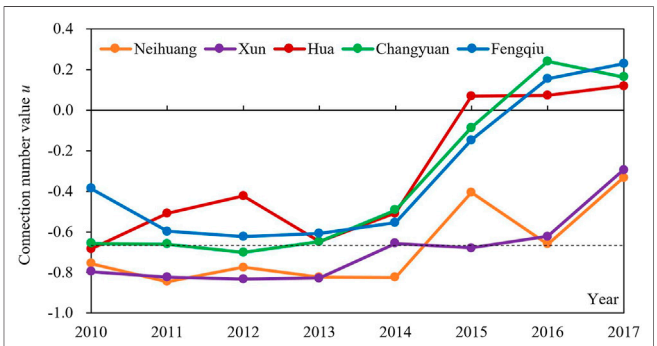


FIGURE 6 | Connection number values of water resources carrying capacity (WRCC) evaluation from 2010 to 2017 in the Dagong irrigation district. The dashed line in the figure represents the boundary of the connection number value for water resources critical overloaded and overloaded status (–0.667) in the irrigation district.

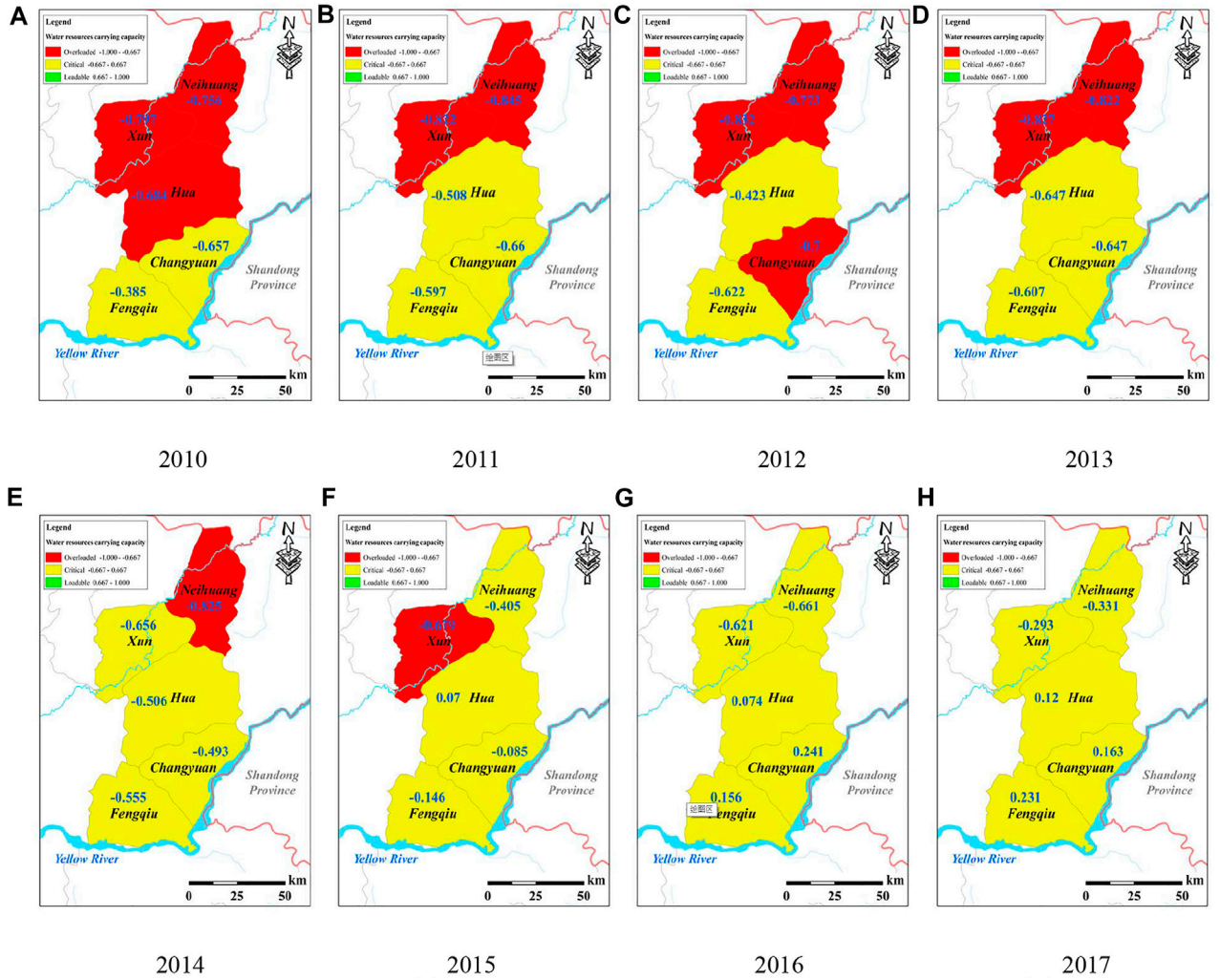


FIGURE 7 | Spatial distributions of water resources carrying capacity (WRCC) evaluation grade values from 2010 to 2017 in the Dagong irrigation district.

0.260 (critical overloaded), respectively. Additionally, in 2016 and 2017, the connection number values of the three counties (Fengqiu County, Changyuan County, and Hua County) were larger than zero and approached loadable status (**Figure 6**). The WRCC in the irrigation district were significantly improved and. it indicates that although the carrying situation in the Dagong irrigation district is severe in recent years, it has developed in a favorable direction as a whole (Zhang et al., 2019; Jin et al., 2021). Thus, there is a need to continuously strengthen the scientific measures of water resources regulation. In addition, the Dagong irrigation district may not have paid enough attention to water resources management before 2013, resulting in the exploitation of water resources far exceeding the available amount. The WRCC system is unbalanced and the WRCC is weak. However, since the implementation of the strictest water resources management system in China (Zuo et al., 2014; Wang et al., 2018), the carrying situation in the irrigation district has been improved markedly, which is consistent with the studies of Du et al. (2020) and Zhang X. et al. (2020).

From the perspective of each county, from 2010 to 2017, the WRCC of the five counties tended to be improved, wherein the magnitudes in Hua County, Changyuan County, and Fengqiu County were significantly higher than Neihuang County and Xun County, especially since 2015. Also, the connection number value in Neihuang County decreased from 2015 to 2016, reflecting the decline of WRCC (**Figure 6**). The average values of the connection number from 2015 to 2017 in Hua County, Changyuan County, and Fengqiu County were 0.088, 0.106, and 0.080, respectively, all greater than zero, while those in Neihuang County and Xun County were −0.466 and −0.531, which were close to the edge of overloaded status. Moreover, in 2017, the order of carrying status in the five counties from excellent to poor was Fengqiu County, Changyuan County, Hua County, Xun County, and Neihuang County, with the corresponding connection number values of 0.231, 0.163, 0.120, −0.293, and −0.331, respectively. It is consistent with the fact that these five counties become farther away from the main stream of the Yellow River (**Figure 7**). Therefore, the WRCC in each county is

significantly related to its distance from the Yellow River (Jin et al., 2021).

In reality, there are many upstream and downstream management units of the backbone projects in the Dagong irrigation district, wherein their relationships are complex and difficult to coordinate, and the upstream water resources supply conditions cannot meet the downstream demand (Zhang X. Y. et al., 2020). Taking Neihuang County at the downstream end of the irrigation canal system as an example, this county needs to apply to the Dagong management office of Xinxiang City and coordinate with the water conservancy bureaus of the upstream counties for water resources use (Song and Zhang, 2018). There are many intermediate links that make it extremely difficult to use water resources. In addition, Neihuang County has only been able to divert water once a year in flood season since 2009, with a time of 1 week. Furthermore, the Yellow River diversion and storage projects are dry most of the time (Zhang, 2017). Combined with the statistical data, in 2017, the available water resources amount per capita in Fengqiu County, Changyuan County, Hua County, Xun County, and Neihuang County are 340.00 m^3, 298.50 m^3, 280.56 m^3, 178.49 m^3, and 278.00 m^3, respectively. Correspondingly, the average water deficient ratios from 2010 to 2017 are 27.91%, 29.91%, 45.60%, 54.89%, and 41.04%, respectively. The above results are also consistent with the studies of Zhang et al. (2015) and Du et al. (2020), indicating that the WRCC of Changyuan County and Fengqiu County have begun to be strengthened since 2014, and the WRCC in the Dagong irrigation district improved in 2017, and the exceptions are Neihuang County and Xun County. Therefore, to promote the WRCC of each county in the Dagong irrigation district to a direction of improvement, this study further analyzes the change of the connection number value for each evaluation index year to year, and diagnoses the main obstacle factors of WRCC.

Affected by many factors such as nature, the economy, society, population, science and technology, the WRCC system in irrigation district is a typical complex system with great uncertainty, mainly including a kind of fuzzy uncertainty due to the uncertainty of the boundary between evaluation samples and evaluation criteria. In this study, SPA is applied to the WRCC evaluation in the Yellow River irrigation district, two sets of evaluation index value and evaluation grade criteria constitute a set pair, and the certainty and uncertainty of their proximity are quantitatively analyzed from three aspects of the same (the evaluation sample value is at the same level as the evaluation level), different (adjacent level) and inverse (separated level). Specifically, SPA divides the certainty of this set pair into similarity and opposition components, which are quantitatively described by a and c, respectively. In addition, the uncertainty between similarity and opposition is divided into the uncertainty at the macro level, measured by b, and that at the micro level, measured by I. Therefore, the calculation method of dynamic difference degree coefficient I varying with evaluation sample proposed in this study, quantifies the uncertainty at the micro level of the set pair constituted by index value and grade criteria. From the perspective of information utilization, this method retains the extremely important information of the variation range of sample data for the research objective. In a word, this study further considers the boundary fuzzy uncertainty of WRCC system in the Yellow River irrigation district, making the evaluation results more comprehensive, more objective, and closer to reality.

Diagnosis of WRCC Obstacle Factors for Five Counties in Irrigation District

The changes of WRCC for each county in recent years were analyzed and discussed above. Then, the obstacle factors of WRCC were diagnosed by the connection number value and semipartial subtraction set pair potential of evaluation index to provide a basis for water resources regulation and control in the Dagong irrigation district.

(1) The connection number value of evaluation index was calculated using the connection number components and difference degree coefficients according to **Eq. 11**. The average values of the connection number for each index from 2010 to 2017 for five counties in the irrigation district are listed in **Table 2**. As shown in **Table 2**, there were six, six, five, two, and three strong obstacle indexes ($-1.0 \le u < -0.6$) in Neihuang County, Xun County, Hua County, Changyuan County, and Fengqiu County, respectively, on average for many years. Nevertheless, there were zero, zero, one, one, and one strong promotion indexes ($0.6 < u \le 1.0$), respectively. These results were consistent with the above results that the WRCC of Fengqiu County, Changyuan County, and Hua County were stronger than Neihuang County and Xun County (**Figures 4**, **7**). It indicates that the numbers of strong obstacle indexes and the strong promotion indexes determined by the connection number value in this study can reflect the overall carrying situation of each county. The more the strong obstacle index and the less the strong promotion index, the worse the carrying status, and vice versa.

Over the years, the utilization ratio of water resources X_2 was a strong obstacle index, available water resources amount per capita X_1, GDP per capita X_4, and water deficient ratio X_6 were strong or middle obstacle indexes for the five counties (**Table 2**). These are the key factors that hinder the improvement of WRCC in the Dagong irrigation district, and also the important objects of WRCC regulation and control. Meanwhile, the annual variations of the connection number value for these indexes are so small that they are difficult to be improved and regulated.

(2) The connection number values of nine WRCC evaluation indexes from 2010 to 2017 for five counties in the Dagong irrigation district, and the semipartial subtraction set pair potential calculated by **Eq. 16**, are shown in **Figure 8**. The change of the connection number value for each index was analyzed and the key influencing factors of carrying status in each county were identified. Moreover, the strong obstacle index and middle obstacle index were the main reasons for weak WRCC, which can be diagnosed as the obstacle factors of WRCC and key objects to be improved.

TABLE 2 | Average values of connection number for water resources carrying capacity (WRCC) evaluation index from 2010 to 2017 in the Dagong irrigation district.

Evaluation index	Neihuang county		Xun county		Hua county		Changyuan county		Fengqiu county	
	Connection number value	Index type	Connection number value	Index type	Connection number value	Index type	Connection number value	Index type	Connection number value	Index type
Available water resources amount per capita X_1 (m^3)	−0.58	Middle obstacle	−0.80	Strong obstacle	−0.80	Strong obstacle	−0.80	Strong obstacle	−0.76	Strong obstacle
Utilization ratio of water resources X_2 (%)	−0.88	Strong obstacle	−0.92	Strong obstacle	−0.90	Strong obstacle	−0.79	Strong obstacle	−0.79	Strong obstacle
Average urbanization ratio X_3 (%)	−0.31	Middle obstacle	0.10	Weak obstacle	0.81	Strong promotion	−0.43	Middle obstacle	0.41	Weak promotion
GDP per capita X_4 (Yuan)	−0.72	Strong obstacle	−0.63	Strong obstacle	−0.75	Strong obstacle	−0.29	Middle obstacle	−0.81	Strong obstacle
Effective irrigation area ratio X_5 (%)	−0.55	Middle obstacle	−0.67	Strong obstacle	0.50	Weak promotion	−0.31	Middle obstacle	0.04	Weak obstacle
Water deficient ratio X_6 (%)	−0.79	Strong obstacle	−0.86	Strong obstacle	−0.81	Strong obstacle	−0.25	Middle obstacle	−0.32	Middle obstacle
Shallow groundwater exploitation ratio X_7 (%)	−0.79	Strong obstacle	−0.46	Middle obstacle	−0.81	Strong obstacle	1.00	Strong promotion	1.00	Strong promotion
Effective utilization coefficient of irrigation water X_8	−0.71	Strong obstacle	−0.34	Middle obstacle	−0.26	Middle obstacle	−0.42	Middle obstacle	−0.40	Middle obstacle
Water consumption ratio of ecological environment X_9 (%)	−0.63	Strong obstacle	−0.91	Strong obstacle	−0.30	Middle obstacle	−0.30	Middle obstacle	−0.34	Middle obstacle

The change trend of the connection number value for each index from 2010 to 2017 was consistent with the semipartial subtraction set pair potential for the five counties, while the overall variation range of the connection number value was larger. However, when the value varied markedly during 2 years, the range of the set pair potential was greater (**Figure 8**). For example, the connection number values of water consumption ratio of ecological environment X_9 for Hua County in 2016 and 2017 were 0.97 and 0.69 (the absolute difference was 0.28), respectively, while the set pair potential were 0.97 and 0.76 (0.21) respectively. In addition, the connection number values of average urbanization ratio X_3 for Fengqiu County in 2013 and 2014 were −0.81 and 0.68 (the absolute difference was 1.49), respectively, while the set pair potential were −0.83 and 0.76 (1.59) respectively. It indicates that the connection number value determined by the calculation method of dynamic difference degree coefficient proposed in this study is reasonable and reliable, and can more sensitively and accurately identify the change of carrying status for evaluation index.

The obstacle factors of WRCC (the connection number value $u < -0.2$) in Neihuang County were X_1, X_2, X_4, X_5, X_6, X_7, and X_8 (**Figure 8A**). It reflected that in recent years, this county was generally in water resources overloaded status because of the less water resources amount, large population, low economic level, high urbanization degree, less effective irrigation area, and low utilization coefficient of irrigation water (Zhang et al., 2017). Furthermore, the reasons for the change in carrying situation were analyzed to guide water resources management, according to the connection number value of index. The improvement of WRCC in 2015 (**Figure 5A**) was mainly due to the water consumption ratio of ecological environment X_9 developed from strong obstacle (the connection number value was −1.00) in 2014 to the strong promotion (0.74) in 2015 (**Figure 8A**). Additionally, the carrying status worsened again in 2016 (**Figure 5A**) because the shallow groundwater exploitation ratio X_7 further deteriorated from strong obstacle (−0.72) in 2015 to the largest strong obstacle (−1.00) in 2016, and meanwhile X_9 deteriorated from strong promotion (0.74) in 2015 to strong obstacle (−0.68) in 2016 (**Figure 8A**). Therefore, the main season for the declined WRCC from 2015 to 2016 is the sharp drop of water resources supply for the ecological environment, followed by the continuous degree increase of groundwater exploitation. Moreover, the reason for the improvement of WRCC in 2017 was that the average urbanization ratio X_3, effective irrigation area ratio X_5, and X_9 were improved (**Figure 5A and 8A**) (Song and Zhang, 2018). Therefore, the key measures to improve the severe water resources carrying situation in Neihuang County are to reasonably control the levels of shallow groundwater exploitation and urban development, and to increase the effective irrigation area and water resources consumption for ecological environment.

The obstacle factors in Xun County were X_1, X_2, X_4, X_6, X_7, and X_9 (**Figure 8**). Furthermore, the main reasons for the

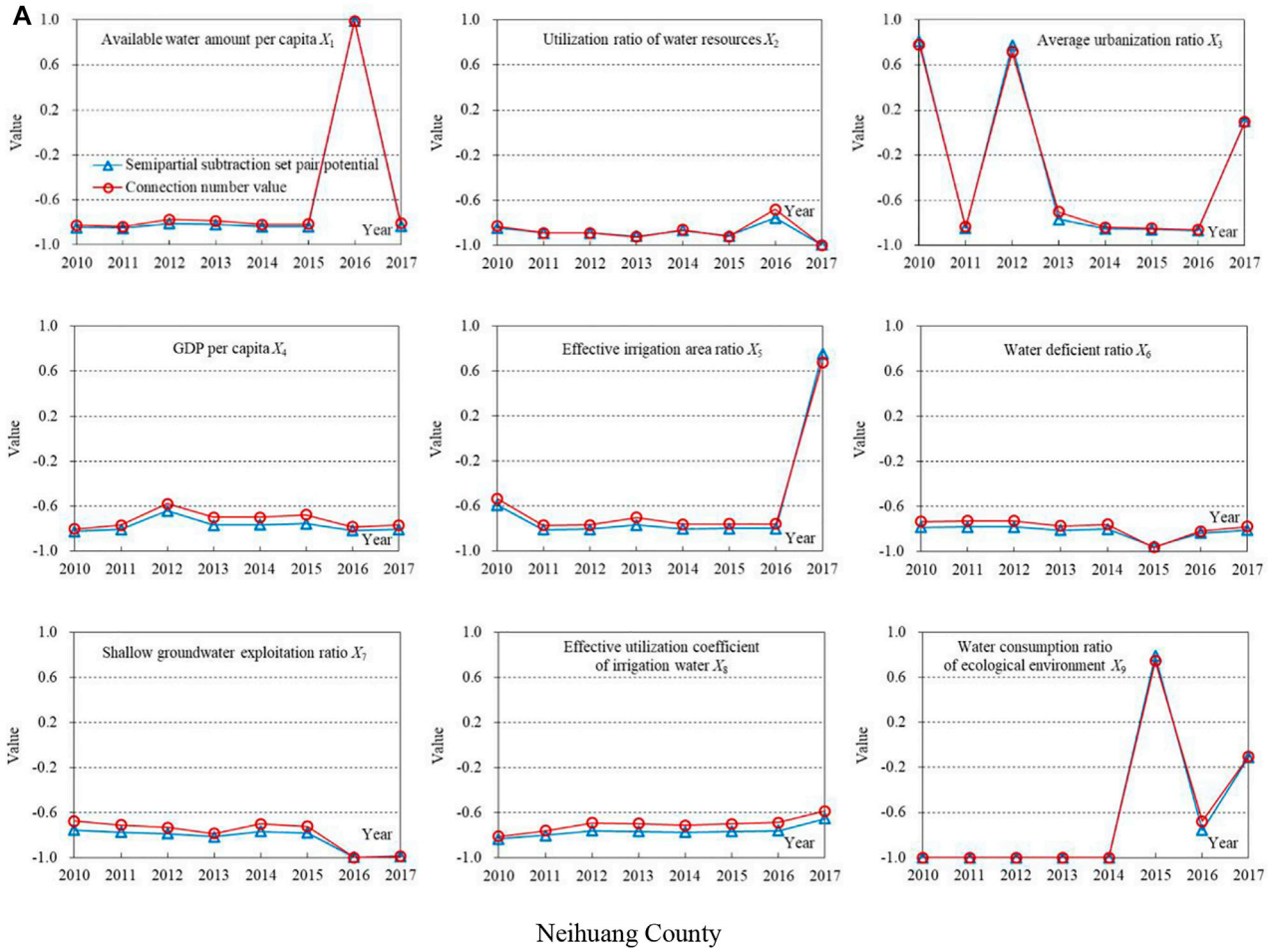


FIGURE 8 | Connection number values of water resources carrying capacity (WRCC) evaluation index from 2010 to 2017 in the Dagong irrigation district. The four dashed lines from top to bottom in the figures represent the boundary of connection number value for strong promotion and weak promotion types (0.6), and those for weak promotion and weak obstacle (0.2), weak obstacle and middle obstacle (–0.2), and middle obstacle and strong obstacle (–0.6) in each county, respectively.

overloaded status from 2010 to 2013 (**Figure 5B**) were that the water resources amount was low, the population was large, the economic level was low, and the effective irrigation area was small (**Figure 8**) (Jin et al., 2021). The slight improvement of WRCC since 2014 (**Figure 5B**) was due to the effective irrigation area ratio X_5 and effective utilization coefficient of irrigation water X_8 gradually developed from strong obstacle to strong promotion. Nevertheless, at the same time, the average urbanization ratio X_3 degenerated from middle obstacle to strong obstacle, and the shallow groundwater exploitation ratio X_7 gradually degenerated from strong promotion to weak obstacle and even strong obstacle (**Figure 8**). It indicated that in recent years, the increase of effective irrigation area had promoted the improvement of WRCC in Xun County. Meanwhile, the rapid development of urbanization and the increasingly high rates of groundwater exploitation had limited the magnitude of improvement, which should be paid attention to. These were consistent with research of Zhang X. et al. (2020), who found that Xun County was in the edge of overloaded status. This may be due to the fact that, for satisfying the water consumption of social and economic development, the expansion of groundwater funnel had not been effectively controlled. In addition, this county is located at the downstream end of the canal system in the Dagong irrigation district (**Figure 7**), and the upstream water supply condition cannot meet the downstream water demand, resulting in the poor WRCC.

The obstacle factors in Hua County were X_1, X_2, X_4, X_6, and X_7 (**Figure 8**). In addition, the approaching overloaded status from 2010 to 2014 (**Figure 5C**) was mainly due to the less water resources amount, large population, low economic level, and high degree of water resources utilization (**Figure 8**) (Zhang, 2017). Therefore, supplementing the amount of water resources, accelerating social and economic development, promoting water saving level, and controlling water resources exploitation are effective means to improve the carrying situation. The WRCC increased significantly from 2015 to 2017 (**Figure 5C**) because of the improvement in the effective utilization coefficient of irrigation water X_8 and water consumption ratio of ecological environment X_9. X_8 gradually developed from strong obstacle before 2015 to weak obstacle and strong promotion thereafter,

while X_9 was improved from strong obstacle in 2014 to strong promotion in 2017 (**Figure 8**). These are directly related to the control of agricultural and ecological water consumption by the policy of "three red lines," and reflect the effectiveness and necessity of the strictest water resources management system in China (Zuo et al., 2014; Zhang et al., 2015). Furthermore, according to the investigation, the amount of water diversion from the Yellow River in Hua County has been about 70 million m^3 since 2007, without taking into account the normal water supply, excessive water resources were stored in reservoirs. This effectively alleviates the problem of farmland irrigation, improves the ecological environment, and guarantees the sustainable and stable development of the social economy in this county. In addition, the average urbanization ratio X_3 remained the index of strong promotion type for a long time (**Figure 8**). The appropriate urbanization level is an important factor to ensure excellent carrying status in Hua County (**Figure 7**).

The obstacle factors in Changyuan County were X_1, X_2, X_3, X_4, and X_6 (**Figure 8**). Furthermore, the carrying status was improved from near overloaded to critical overloaded from 2014 to 2017 (**Figure 5D**). This was mainly because the effective irrigation area ratio X_5 gradually developed from strong obstacle in 2014 to weak promotion and strong promotion. Meanwhile, the effective utilization coefficient of irrigation water X_8 and water consumption ratio of ecological environment X_9 were improved from strong obstacle to strong promotion (**Figure 8**). It shows that since the implementation of the strictest water resources management system, this county has scientifically adjusted the agricultural and ecological water consumption, and the effects are remarkable (Dou and Wang, 2017; Zhang et al., 2017). However, the connection number values of average urbanization ratio X_3, GDP per capita X_4, and water deficient ratio X_6 fluctuated markedly over the years, and they all belonged to the index of middle obstacle type (**Figure 8** and **Table 2**), which should be paid close attention as obstacle factors. The shallow groundwater exploitation ratio X_7 was always strong promotion (**Figure 8**), indicating that Changyuan County has strictly controlled the degree of groundwater exploitation for a long time (Song et al., 2020). This is a crucial factor for guaranteeing its relatively excellent water resources carrying situation (**Figure 7**).

The obstacle factors in Fengqiu County were X_1, X_2, X_4, and X_6 (**Figure 8**). It can be seen that the less water resources amount, large population, and low economic level significantly restricted the improvement of its WRCC (Zhang, 2017; Jin et al., 2021). Therefore, transferring and supplementing the water resources and promoting the economic development are powerful measures to strengthen the WRCC. Similar to Changyuan County, Fengqiu County gradually developed from near overloaded status to critical overloaded status from 2014 to 2017 (**Figure 5e**), which was mainly due to that the effective irrigation area ratio X_5, effective utilization coefficient of irrigation water X_8, and water consumption ratio of ecological environment X_9 were improved from middle obstacle or strong obstacle to strong promotion (**Figure 8**). It suggests that this county has strictly carried out the control of "red lines" on agricultural and ecological water consumption (Zhang et al., 2015), fully ensuring the effective improvement of its carrying situation. Moreover, the average urbanization ratio X_3 quickly recovered to weak promotion and strong promotion after deteriorating to strong obstacle in 2013, while the water deficient ratio X_6 was not improved after worsening to strong obstacle in 2012 from strong promotion in 2010 and 2011. The shallow groundwater exploitation ratio X_7 was the index of strong promotion type for a long time (**Figure 8**) (Song et al., 2020). It reflects that the appropriate levels of groundwater exploitation and urbanization are key factors to ensure the strong WRCC in Fengqiu County (**Figure 7**). Meanwhile, it is necessary to take measures, such as adding water supply, decreasing water consumption, or increasing water use efficiency, to reduce the water deficient ratio and then strengthen the WRCC.

In summary, the evaluation and diagnosis results obtained in this study are consistent with the real water resources carrying situation in the Dagong irrigation district under the current policy environment (Xu, 2020; Fan et al., 2021; Qin, 2021). According to the construction standard of ecological irrigation district, and great concerns about ecological civilization construction and new water resources management concept in China, the administrative departments should further improve the water resources allocation of each county in the irrigation district and conduct other relevant works. So as to realize that there are water resources for supplementing source in flood season and those for irrigation in non-flood season, on the basis of guaranteeing the domestic and industrial water supply.

CONCLUSION

In order to deal with the uncertainty between evaluation samples and evaluation criteria, a connection number model of WRCC evaluation was established. Furthermore, taking account of the information carried by sample data, a calculation method of dynamic difference degree coefficient varying with the actual evaluation sample was also proposed. As a result, a quantitative evaluation and obstacle factor diagnosis model of WRCC in the Yellow River irrigation district was constructed. In addition, an empirical study was carried out in the Dagong ecological irrigation district, and the following main conclusions were obtained:

1) The results of WRCC evaluation and diagnosis in a typical irrigation district were consistent with the real situation and existing studies, indicating that the connection number evaluation and diagnosis method based on dynamic difference degree coefficient was effective and reliable. These results provided important scientific support for water resources allocation and management in the Yellow River irrigation district.
2) Although the WRCC in the Dagong irrigation district remained severe, it had been improved as a whole since 2013. Five counties of the irrigation district were all in water resources critical overloaded status by 2017. Moreover, the WRCC of Fengqiu County, Changyuan County, Hua County, Xun County, and Neihuang County

became weaker in turn, which was consistent with the fact that they were further away from the main stream of the Yellow River.
3) The utilization ratio of water resources, available water resources amount per capita, GDP per capita, and water deficient ratio belonged to the index of strong or middle obstacle type. These were the main obstacle factors of WRCC, as well as the key and difficulty of water resources regulation and control for the Dagong irrigation district.
4) The connection number model based on dynamic difference degree coefficient proposed in this study can identify the relatively certain water resources carrying situation and its crucial obstacle factors in irrigation district, and can be applied to resources, environment, ecology carrying capacity fields, and other set pair system problems.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

YC prepared, created, and presented the published work; applied statistical, mathematical, and computational technique to analyze or synthesize study data. LZ, LJ, and YC conceptualized and formulated overarching research goals and aims; developed and designed the methodology; and was involved in project administration, and funding acquisition. GW, BZ, and WN curated, visualized, and presented the data. All authors were involved in critical review, commentary, and revision.

FUNDING

This study was financially supported by the National Natural Science Foundation of China (Grant No. 52109009) and National Key Research and Development Program of China (Grant No. 2018YFC0407206), the Natural Science Foundation of Anhui Province, China (Grant No. 2108085QE254), and the Fundamental Research Funds for the Central Universities (Grant Nos. JZ2021HGTA0165, JZ2020HGQA0202, JZ2021HGQB0281).

ACKNOWLEDGMENTS

The authors would like to thank the handling editor and the reviewers, for their reviews and valuable comments that significantly improved the quality of this paper.

REFERENCES

Chang, T., Li, J. Q., Jin, J. L., Chen, L., Dong, T., Chen, M. L., et al. (2020). Multi-dimensional Water Resources Carrying Capacity Evaluation index System for Water Flow System Function. *Water Resour. Prot.* 36, 44–51. doi:10.3880/j.issn.1004-6933.2020.01.007

Chen, S. Y. (2005). *Theories and Methods of Variable Fuzzy Sets in Water Resources and Flood Control System*. Dalian: Dalian University of Technology Press.

Chong, T., Yi, S., and Heng, C. (2017). Application of Set Pair Analysis Method on Occupational hazard of Coal Mining. *Saf. Sci.* 92, 10–16. doi:10.1016/j.ssci.2016.09.005

Cui, Y., Feng, P., Jin, J., and Liu, L. (2018). Water Resources Carrying Capacity Evaluation and Diagnosis Based on Set Pair Analysis and Improved the Entropy Weight Method. *Entropy* 20, 359. doi:10.3390/e20050359

Dai, D., Sun, M., Xu, X., and Lei, K. (2019). Assessment of the Water Resource Carrying Capacity Based on the Ecological Footprint: a Case Study in Zhangjiakou City, North China. *Environ. Sci. Pollut. Res.* 26, 11000–11011. doi:10.1007/s11356-019-04414-9

Dou, M., and Wang, Y. (2017). The Construction of a Water Rights System in China that Is Suited to the Strictest Water Resources Management System. *Water Supply* 17, 238–245. doi:10.2166/ws.2016.130

Du, X. F., Li, Y. B., and Zhang, X. Y. (2020). Study on Water Resources Carrying Capacity of Ecological Diversion Irrigation District in the Lower Reaches of the Yellow River. *Hydro-Science Eng.* (2), 22–29. doi:10.12170/20200209001

Fan, X. C., Qin, J. T., Xu, L., Liu, S. R., Gu, S. W., and Lyu, M. C. (2021). Construction and Empirical Analysis of the Evaluation index System for the Water-Saving Level of Large-Sized Irrigation Districts. *Trans. Chin. Soc. Agric. Eng.* 37, 99–107. doi:10.11975/j.issn.1002-6819.2021.20.011

Song, F., Yang, X., and Wu, F. (2020). Catastrophe Progression Method Based on M-K Test and Correlation Analysis for Assessing Water Resources Carrying Capacity in Hubei Province. *J. Water Clim. Change* 11, 556–567. doi:10.2166/wcc.2018.114

Gonçalves, J. M., Pereira, L. S., Fang, S. X., and Dong, B. (2007). Modelling and Multicriteria Analysis of Water Saving Scenarios for an Irrigation District in the Upper Yellow River Basin. *Agric. Water Manage.* 94, 93–108. doi:10.1016/j.agwat.2007.08.011

Gong, L., and Jin, C. (2009). Fuzzy Comprehensive Evaluation for Carrying Capacity of Regional Water Resources. *Water Resour. Manage.* 23, 2505–2513. doi:10.1007/s11269-008-9393-y

He, L., Du, Y., Wu, S., and Zhang, Z. (2021). Evaluation of the Agricultural Water Resource Carrying Capacity and Optimization of a Planting-Raising Structure. *Agric. Water Manage.* 243, 106456. doi:10.1016/j.agwat.2020.106456

Jia, Z., Wu, Z., Luo, W., Xi, W., Tang, S., Liu, W. L., et al. (2013). The Impact of Improving Irrigation Efficiency on Wetland Distribution in an Agricultural Landscape in the Upper Reaches of the Yellow River in China. *Agric. Water Manage.* 121, 54–61. doi:10.1016/j.agwat.2013.01.003

Jin, J. L., Wei, Y. M., and Ding, J. (2004). Fuzzy Comprehensive Evaluation Model Based on Improved Analytic Hierarchy Process. *J. Hydraulic Eng.* 35 (3), 65–70. doi:10.13243/j.cnki.slxb.2004.03.011

Jin, J. L., Wu, K. Y., and Wei, Y. M. (2008). Connection Number Based Assessment Model for Watershed Water Security. *J. Hydraulic Eng.* 39, 401–409. doi:10.3321/j.issn:0559-9350.2008.04.003

Jin, J. L., Dong, T., Li, J. Q., Zhang, L. B., and Li, H. (2018). Water Resources Carrying Capacity Evaluation Method under Different Carrying Standards. *Adv. Water Sci.* 29, 31–39. doi:10.14042/j.cnki.32.1309.2018.01.004

Jin, J. L., Zhang, H. Y., Chen, M. L., Cui, Y., and Ning, S. W. (2019a). Evaluation and Diagnosis of Agricultural Drought Vulnerability Based on Grey Correlation and Connection Number Coupling. *J. Catastrophol.* 34, 1–7. doi:10.3969/j.issn.1000-811X.2019.01.001

Jin, J. L., Zhang, H. Y., Chen, M. L., Wang, H. R., Cui, Y., and Ning, S. W. (2019b). Evaluation of Regional Agricultural Drought Vulnerability Coupled with Dynamic Connection Number and Full Partial Certainty. *J. Beijing Normal Univ. (Nat. Sci.)* 55, 724–730. doi:10.16360/j.cnki.jbnuns.2019.06.008

Jin, J. L., Shen, S. X., Cui, Y., Zhang, X. Y., He, P., and Ning, S. W. (2021). Dynamic Evaluation of Water Resources Carrying Capacity in the Yellow River Diversion Irrigation District Based on Semipartial Subtraction Set Pair Potential. 52 507–520. doi:10.13243/j.cnki.slxb.20200561

Kang, J., Zi, X., Wang, S., and He, L. (2019). Evaluation and Optimization of Agricultural Water Resources Carrying Capacity in Haihe River Basin, China. *Water* 11, 999. doi:10.3390/w11050999

Kumar, K., and Garg, H. (2016). TOPSIS Method Based on the Connection Number of Set Pair Analysis under Interval-Valued Intuitionistic Fuzzy Set Environment. *Comp. Appl. Math.* 37, 1319–1329. doi:10.1007/s40314-016-0402-0

Li, Z. Y., Wu, M., Liu, Z. Y., Li, D. P., and Guo, C. (2009). A New Approach to *I* in Connection Number and Application in Water Quality Assessment. *J. Sichuan Univ. (Engineer. Sci. Ed.)* 41, 8–13. doi:10.15961/j.jsuese.2009.01.002

Li, M., Zheng, T., Zhang, J., Fang, Y., Liu, J., Zheng, X., et al. (2019). A New Risk Assessment System Based on Set Pair Analysis - Variable Fuzzy Sets for Underground Reservoirs. *Water Resour. Manage.* 33, 4997–5014. doi:10.1007/s11269-019-02390-w

Li, Z., Jin, J., Cui, Y., Zhang, L., Wu, C., Ning, S., et al. (2021). Dynamic Evaluation of Regional Water Resources Carrying Capacity Based on Set Pair Analysis and Partial Connection Number. *J. Hydraulic Eng.*. doi:10.2166/ws.2021.371

Liao, X., Ren, Y., Shen, L., Shu, T., He, H., and Wang, J. (2020). A "Carrier-Load" Perspective Method for Investigating Regional Water Resource Carrying Capacity. *J. Clean. Prod.* 269, 122043. doi:10.1016/j.jclepro.2020.122043

Lyu, H.-M., Shen, S.-L., and Zhou, A. (2021). The Development of IFN-SPA: A New Risk Assessment Method of Urban Water Quality and its Application in Shanghai. *J. Clean. Prod.* 282, 124542. doi:10.1016/j.jclepro.2020.124542

Miao, Q., Shi, H., Gonçalves, J. M., and Pereira, L. S. (2015). Field Assessment of basin Irrigation Performance and Water Saving in Hetao, Yellow River basin: Issues to Support Irrigation Systems Modernisation. *Biosyst. Eng.* 136, 102–116. doi:10.1016/j.biosystemseng.2015.05.010

Pan, Z. W., Jin, J. L., and Zhou, R. X. (2016). Assessment Method Using Connection Number Based on Trapezoidal Fuzzy Numbers and its Application. *Adv. Sci. Technol. Water Resour.* 36, 69–74. doi:10.3880/j.issn.1006-7647.2016.05.013

Pan, Z., Wang, Y., Jin, J., and Liu, X. (2017). Set Pair Analysis Method for Coordination Evaluation in Water Resources Utilizing Conflict. *Phys. Chem. Earth, Parts A/B/C* 101, 149–156. doi:10.1016/j.pce.2017.05.009

Peng, T., and Deng, H. (2020). Comprehensive Evaluation on Water Resource Carrying Capacity Based on DPESBR Framework: A Case Study in Guiyang, Southwest China. *J. Clean. Prod.* 268, 122235. doi:10.1016/j.jclepro.2020.122235

Peng, T., Deng, H., Lin, Y., and Jin, Z. (2021). Assessment on Water Resources Carrying Capacity in Karst Areas by Using an Innovative DPESBRM Concept Model and Cloud Model. *Sci. Total Environ.* 767, 144353. doi:10.1016/j.scitotenv.2020.144353

Pereira, L. S., Gonçalves, J. M., Dong, B., Mao, Z., and Fang, S. X. (2007). Assessing basin Irrigation and Scheduling Strategies for Saving Irrigation Water and Controlling Salinity in the Upper Yellow River Basin, China. *Agric. Water Manage.* 93, 109–122. doi:10.1016/j.agwat.2007.07.004

Qi, P., Xia, Z., Zhang, G., Zhang, W., and Chang, Z. (2021). Effects of Climate Change on Agricultural Water Resource Carrying Capacity in a High-Latitude basin. *J. Hydrol.* 597, 126328. doi:10.1016/j.jhydrol.2021.126328

Qin, T. (2021). *Comprehensive Evaluation of Water Resources Security in Dagong Yellow River Diversion Irrigation Area of Henan Province*. Zhengzhou: North China University of Water resources and Electric Power.

Ren, D., Xu, X., Engel, B., and Huang, G. (2018). Growth Responses of Crops and Natural Vegetation to Irrigation and Water Table Changes in an Agro-Ecosystem of Hetao, Upper Yellow River basin: Scenario Analysis on maize, sunflower, Watermelon and Tamarisk. *Agric. Water Manage.* 199, 93–104. doi:10.1016/j.agwat.2017.12.021

Roy, D. K., and Datta, B. (2019). Adaptive Management of Coastal Aquifers Using Entropy-Set Pair Analysis-Based Three-Dimensional Sequential Monitoring Network Design. *J. Hydrol. Eng.* 24, 04018072. doi:10.1061/(ASCE)HE.1943-5584.0001765

Song, G. X., and Yang, D. L. (2003). Methods for Identifying and Improving the Consistency of Fuzzy Judgment Matrix. *Syst. Eng.* 21, 110–116.

Song, D. J., and Zhang, L. X. (2018). Investigation and Suggestions on the Current Situation of Dagong Yellow River Diversion Irrigation District. *Henan Water Resour. South-to-North Water Diversion* 47 (5), 25–26.

Song, X.-m., Kong, F.-z., and Zhan, C.-s. (2011). Assessment of Water Resources Carrying Capacity in Tianjin City of China. *Water Resour. Manage.* 25, 857–873. doi:10.1007/s11269-010-9730-9

Song, Q. X., Li, J., Miao, H. C., Gao, Y. Y., Wang, Z. P., and Chen, F. (2020). Analysis of Groundwater Exploitation Potential in the north Region of Yellow River in Henan Province. *Yellow River* 42, 67–72. doi:10.3969/j.issn.1000-1379.2020.08.014

Su, Y., Gao, W., and Guan, D. (2019). Integrated Assessment and Scenarios Simulation of Water Security System in Japan. *Sci. Total Environ.* 671, 1269–1281. doi:10.1016/j.scitotenv.2019.03.373

Tang, Y. P. (2009). The Methods to Fetch *I* in Difference Degree Coefficient of Set Pair Analysis and its Applications. *Math. Pract. Theor.* 39, 67–70.

Wan, X., Yang, T., Zhang, Q., Yan, X., Hu, C., Sun, L., et al. (2021). A Novel Comprehensive Model of Set Pair Analysis with Extenics for River Health Evaluation and Prediction of Semi-arid basin - A Case Study of Wei River Basin, China. *Sci. Total Environ.* 775, 145845. doi:10.1016/j.scitotenv.2021.145845

Wang, W., Jin, J., Ding, J., and Li, Y. (2009). A New Approach to Water Resources System Assessment - Set Pair Analysis Method. *Sci. China Ser. E-technol. Sci.* 52, 3017–3023. doi:10.1007/s11431-009-0099-z

Wang, S., Yang, F.-L., Xu, L., and Du, J. (2013). Multi-scale Analysis of the Water Resources Carrying Capacity of the Liaohe Basin Based on Ecological Footprints. *J. Clean. Prod.* 53, 158–166. doi:10.1016/j.jclepro.2013.03.052

Wang, Y., Wang, Y., Su, X., Qi, L., and Liu, M. (2019). Evaluation of the Comprehensive Carrying Capacity of Interprovincial Water Resources in China and the Spatial Effect. *J. Hydrol.* 575, 794–809. doi:10.1016/j.jhydrol.2019.05.076

Wu, C., Zhou, L., Jin, J., Ning, S., Zhang, Z., and Bai, L. (2020). Regional Water Resource Carrying Capacity Evaluation Based on Multi-Dimensional Precondition Cloud and Risk Matrix Coupling Model. *Sci. Total Environ.* 710, 136324. doi:10.1016/j.scitotenv.2019.136324

Wang, X.-j., Zhang, J.-y., Gao, J., Shahid, S., Xia, X.-h., Geng, Z., et al. (2018). The New Concept of Water Resources Management in China: Ensuring Water Security in Changing Environment. *Environ. Dev. Sustain.* 20, 897–909. doi:10.1007/s10668-017-9918-8

Xiong, L., Xu, X., Engel, B., Huang, Q., Huo, Z., Xiong, Y., et al. (2021). Modeling Agro-Hydrological Processes and Analyzing Water Use in a Super-large Irrigation District (Hetao) of Arid Upper Yellow River basin. *J. Hydrol.* 603, 127014. doi:10.1016/j.jhydrol.2021.127014

Xu, M. J. (2020). *Analysis and Prediction of Cultivated Land Grain Efficiency in Typical Yellow River Diversion Irrigation Area of Henan Province*. Tianjin: Tiangong University.

Yang, X.-H., He, J., Di, C.-L., and Li, J.-Q. (2014). Vulnerability of Assessing Water Resources by the Improved Set Pair Analysis. *Therm. Sci.* 18, 1531–1535. doi:10.2298/TSCI1405531Y

Yang, J., Lei, K., Khu, S., and Meng, W. (2015). Assessment of Water Resources Carrying Capacity for Sustainable Development Based on a System Dynamics Model: A Case Study of Tieling City, China. *Water Resour. Manage.* 29, 885–899. doi:10.1007/s11269-014-0849-y

Yin, Z., Ottlé, C., Ciais, P., Zhou, F., Wang, X., Jan, P., et al. (2021). Irrigation, Damming, and Streamflow Fluctuations of the Yellow River. *Hydrol. Earth Syst. Sci.* 25, 1133–1150. doi:10.5194/hess-25-1133-2021

Wang, Y., Cheng, H., and Huang, L. (2018). Water Resources Carrying Capacity Evaluation of a Dense City Group: A Comprehensive Water Resources Carrying Capacity Evaluation Model of Wuhan Urban Agglomeration. *Urban Water J.* 15, 615–625. doi:10.1080/1573062X.2018.1529805

Zhang, Y. F., Guo, W., Xu, J. X., and Zhang, X. Y. (2015). Optimal Allocation to Water Resources of the Yellow River Irrigation District of Dagong Based on the Strictest Managerical System of Water Resources. *J. North China Univ. Water Resour. Electric Power (Natural Sci. Ed.)* 36, 28–32. doi:10.3969/j.issn.1002-5634.2015.03.007

Zhang, S., Xiang, M., Yang, J., Fan, W., and Yi, Y. (2019). Distributed Hierarchical Evaluation and Carrying Capacity Models for Water Resources Based on Optimal Water Cycle Theory. *Ecol. Indicators* 101, 432–443. doi:10.1016/j.ecolind.2019.01.048

Zhang, Q., Xu, P., Chen, J., Qian, H., Qu, W., and Liu, R. (2021). Evaluation of Groundwater Quality Using an Integrated Approach of Set Pair Analysis and Variable Fuzzy Improved Model with Binary Semantic Analysis: A Case Study in Jiaokou Irrigation District, East of Guanzhong Basin, China. *Sci. Total Environ.* 767, 145247. doi:10.1016/j.scitotenv.2021.145247

Zhang, J. (2017a). *Study on Coordinated Development of Typical Area Irrigated by the Yellow River Based on Sustainable Theory*. Zhengzhou: North China University of Water resources and Electric Power.

Zhang, X., Du, X., and Li, Y. (2020). Comprehensive Evaluation of Water Resources Carrying Capacity in Ecological Irrigation Districts Based on Fuzzy Set Pair Analysis. *Dwt* 187, 63–69. doi:10.5004/dwt.2020.25287

Zhang, X. Y., Du, X. F., Xu, J. X., and Zhang, Z. Q. (2019). Analysis of Water Resources Carrying Capacity in Dagong Yellow River Diversion Irrigation Area Based on Ecology. *Yellow River* 41, 49–52+57. doi:10.3969/j.issn.1000-1379.2019.06.011

Zhang, X. Y., Qin, T., Yang, Q. X., Du, X. F., Yang, L. B., and Liu, X. M. (2020). Method of Evaluating Safety for Water Usage and its Application to Water Use in Irrigation Districts in the Lower Reach of the Yellow River. *J. Irrig. Drain.* 39, 18–24. doi:10.13522/j.cnki.ggps.2020184

Zhang, Y. S., Lu, Z. G., Wang, M., Qin, H. X., Wang, Y. P., He, G., et al. (2017). Calculation and Analysis of Irrigation Water Use Efficiency of Henan Province. *China Rural Water Hydropower* (1), 9–12+17.

Zhao, Y., Wang, Y., and Wang, Y. (2021). Comprehensive Evaluation and Influencing Factors of Urban Agglomeration Water Resources Carrying Capacity. *J. Clean. Prod.* 288, 125097. doi:10.1016/j.jclepro.2020.125097

Zhao, K. Q. (2000). *Set Pair Analysis and its Preliminary Application*. Hangzhou: Zhejiang Science and Technology Press.

Zhou, R., Jin, J., Cui, Y., Ning, S., Bai, X., Zhang, L., et al. (2022). Agricultural Drought Vulnerability Assessment and Diagnosis Based on Entropy Fuzzy Pattern Recognition and Subtraction Set Pair Potential. *Alexandria Eng. J.* 61, 51–63. doi:10.1016/j.aej.2021.04.090

Zuo, Q., Jin, R., Ma, J., and Cui, G. (2014). China Pursues a Strict Water Resources Management System. *Environ. Earth Sci.* 72, 2219–2222. doi:10.1007/s12665-014-3369-4

Zuo, Q., Guo, J., Ma, J., Cui, G., Yang, R., and Yu, L. (2021). Assessment of Regional-Scale Water Resources Carrying Capacity Based on Fuzzy Multiple Attribute Decision-Making and Scenario Simulation. *Ecol. Indicators* 130, 108034. doi:10.1016/j.ecolind.2021.108034

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Environmental Science

ORIGINAL RESEARCH
published: 31 March 2022
doi: 10.3389/fenvs.2022.820615

Spatial–Temporal Pattern and Influencing Factors of Drought Impacts on Agriculture in China

Xiyuan Deng[1,2], Guoqing Wang[1,2,3], Haofang Yan[4], Jintao Zheng[1] and Xuegang Li[1]*

[1]State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Nanjing Hydraulic Research Institute, Nanjing, China, [2]Research Center for Climate Change, Nanjing, China, [3]Yangtze Institute for Conservation and Development, Nanjing, China, [4]Research Centre of Fluid Machinery Engineering and Technology, Jiangsu University, Zhenjiang, China

Agricultural drought disaster is a major natural disaster affecting economic and social development. It is of significance to investigate the spatial–temporal pattern and the dominant influence of natural and human factors on agricultural drought disasters for drought hazard relief. In this study, Mann–Kendall test was adopted to explore the evolution of agricultural drought disasters. Random forest algorithm, which integrates feature importance and accumulated local effects plot, was applied to quantify the effect of principal influencing factors on agricultural drought disasters. Results show that over the period from 1950 to 2019, agricultural drought disasters in China have undergone significant fluctuations. The spatial pattern of agricultural drought disaster tends to decrease in severity from north to south. The total sown area of crops, precipitation, effective irrigation area, domestic patent application authorization, and regional GDP are the top 5 dominant factors influencing agricultural drought disasters. It also found that agricultural drought disaster negatively correlates with precipitation, domestic patent application authorization, and regional GDP, and the nonlinear response of agricultural drought disaster to total sown area of crops and effective irrigation area can be basically divided into two stages. In the first stage, with the increase of feature value, agricultural drought disaster is also increasing. In the second stage, with the increase of feature value, agricultural drought disaster is growing slow or just decreasing. The results can deepen the understanding of agricultural drought disasters and provide scientific basis for drought event monitoring, evaluation, and early warning.

Keywords: agricultural drought disaster, drought impacts, spatial–temporal pattern, influencing factors, human activities, random forest, feature importance, ALE plots

OPEN ACCESS

Edited by:
Zhenzhong Zeng,
Southern University of Science and Technology, China

Reviewed by:
Shanlei Sun,
Nanjing University of Information Science and Technology, China
Jianfeng Li,
Hong Kong Baptist University, Hong Kong SAR, China

****Correspondence:***
Guoqing Wang
gqwang@nhri.cn

Specialty section:
This article was submitted to Environmental Economics and Management, a section of the journal Frontiers in Environmental Science

Received: *23 November 2021*
Accepted: *03 February 2022*
Published: *31 March 2022*

Citation:
Deng X, Wang G, Yan H, Zheng J and Li X (2022) Spatial–Temporal Pattern and Influencing Factors of Drought Impacts on Agriculture in China. Front. Environ. Sci. 10:820615. doi: 10.3389/fenvs.2022.820615

1 INTRODUCTION

Drought is one of the most frequent natural hazards in the world (Schwalm et al., 2017). Compared with other natural disasters, drought usually affects vast space with long period, and has great impact on hydrology and ecosystem (Orth and Destouni, 2018). Drought may bring about the reduction of soil moisture, river runoff, and crop yield; cause the degradation of river ecological function; and even affect regional water, food, and ecological security. Drought is a major natural disaster affecting economic and social development in China (Lv et al., 2011). From 2008 to 2018, the average area affected by drought in China was 12.8 million hectares, with an average grain loss of 18.7 billion kg (Ministry of Water Resources of China, 2019). In the context of global warming, agriculture

production is affected by increasingly serious meteorological disasters. It is quite important to carry out the research on spatial–temporal pattern and evolution characteristics of agricultural drought disasters in China, so as to reduce the potential risk of drought hazard and formulate corresponding measures (Yan et al., 2016; Dayal et al., 2018; Ni et al., 2019; Wu et al., 2020; Zhao et al., 2020).

The selection of agricultural drought indicators is the basis of agricultural drought analysis. Agricultural drought involves atmosphere, soil, and crops. The commonly used indicators include standardized precipitation index (SPI), standardized soil moisture index (SSMI), water deficit index (WDI), crop moisture index (CMI), and Palmer Drought Severity Index (PDSI) (Cancelliere et al., 2006; Shah and Mishra, 2020; Tian et al., 2022). Most of the aforementioned agricultural indicators are composed of data of crop and soil (Hao et al., 2017; West et al., 2019). They mainly consider disaster drivers and disaster environment, without involving the actual loss of crops directly after drought events. Effective agricultural drought indicators should reflect the impacts of drought (Liu et al., 2019). Li et al. (2019) analyzed the temporal and spatial characteristics of drought in Heilongjiang, Jilin, and Liaoning provinces based on the statistical data of crop sown area, drought-affected area, drought-suffering area, area with no harvest, and grain loss due to severe drought in Northeast China from 1949 to 2017. Wu et al. (2018) adopted drought-affected area rate as drought risk index to establish the risk assessment and zoning model for regional drought disaster based on Cloud Model and Bootstrap Method (CMBM) in Anhui province.

Drought is a natural phenomenon, but it is also widely affected by human activities. For instance, hydrological drought may occur downstream of a reservoir; diversion irrigation may alleviate agricultural drought; long-term overexploitation of groundwater for domestic and production water use may cause ecological drought. Nowadays, human activities have deeply affected water cycle and energy cycle, and the influencing factors of agricultural drought disasters are becoming more and more complex. It is necessary to identify the influencing factors of agricultural drought disasters, screen key disaster-causing factors, and explore the relationship between agricultural drought disaster and various factors, to prevent and control agricultural drought disasters (Yang et al., 2020). Many researches have been carried out on driving factors of agricultural drought disasters (Blauhut et al., 2016; Huang et al., 2015; Pang et al., 2019; Zobeidi et al., 2021). Han et al. (2021) analyzed the influencing factors of agricultural drought in Loess Plateau (LP), and the result shows that significant increasing precipitation ($p < 0.01$) in the LP has not alleviated agricultural drought, whereas significant increasing temperature ($p < 0.01$) is the direct factor inducing agricultural drought, and the implementation of vegetation restoration project further aggravates the risk of agricultural drought. Hong (2017) calculated the information transmission direction and intensity between different types of drought indicators in Hanjiang River Basin, and analyzed the regular pattern of drought propagation and evolution; the result shows that the interaction between agricultural drought and hydrological drought is quite close, showing a high degree of synchronization. However, most of the previous studies on drought are conducted for a river basin or a region rather than the entire China (Huang et al., 2015; Cheng et al., 2017; Wu et al., 2017; Dai et al., 2020), and influencing factors only focus on natural attributes, with less consideration of the role of human activities on drought disasters (Javadinejad et al., 2020).

The response of hydrological and agricultural systems to meteorological conditions is nonlinear (Berghuijs et al., 2016; Konapala and Mishra, 2016); it is one of the sticky points to separate the role of individual natural or human factors and clarify their driving mechanism on agricultural drought disasters. As an algorithm that uses computers to imitate human learning, machine learning can develop learning strategies, analyze potential patterns, and predict target variables according to existing data (Kohavi and Provost, 1998; Liu et al., 2019; Apley and Zhu, 2016). In water cycle field, machine learning algorithm can capture the nonlinear relationship between input variables (e.g., precipitation, temperature) and output variables (e.g., runoff), and apply the functional relationship to target prediction (Nourani et al., 2014; Raghavendra and Deka, 2014). Although most machine learning algorithms cannot directly quantify the internal mechanism of the model behavior (Gupta and Nearing, 2015; Karpatne et al., 2017), the emergence of interpretable methods can improve the understanding of specific machine learning model or prediction (Guidotti et al., 2019; Ji et al., 2019; Liu. W et al., 2019). Based on interpretable methods, it is easier to quantify feature importance, and clarify the dependency between input features and output targets. Nowadays, machine learning algorithms have been widely used in drought prediction (Liu et al., 2018; Başakın et al., 2019; Shamshirband et al., 2020); interpretable methods also began to rise in hydro-meteorological area (Schwalm et al., 2017; Fienen et al., 2018; Koch et al., 2019), but the application of the aforementioned emerging technologies in the identification of key factors of agricultural drought disasters is still limited.

The previous discussion suggests that limited research has been conducted to investigate the dominant nonlinear influence of natural and human factors on agricultural drought disasters over China based on machine learning model. Therefore, the main objectives of this paper are

1) To explore the spatial–temporal pattern and evolution characteristics of agricultural drought disasters in China;
2) To identify the dominant natural and human factors and their effects on agricultural drought disasters in China based on machine learning model.

To achieve these aims, first, we selected drought-affected area and drought-suffering area as the indexes indicating agricultural drought disasters, and the spatial–temporal pattern and evolution characteristics of agricultural drought disasters were presented based on Mann–Kendall test. Second, an index system that consisted of natural and human factors of agricultural drought disasters with 22 indexes was established, and Random Forest algorithm derived from 4 datasets of 2011–2019, 2004–2019, 1987–2019, and 1979–2019 were trained and tested to explore

TABLE 1 | List of indicators; their data sources and brief description are provided

Indicator	Data source	Description	Unit	Time period
Drought-affected area	China Statistical Yearbook	The sown area with crop yield 10% or more lower than normal year due to drought disaster	10^3 ha	1950–2020 (National scale)
Drought-suffering area	China Statistical Yearbook	The sown area with crop yield 30% or more lower than normal year due to drought disaster	10^3 ha	1979–2019 (Provincial scale)
Total sown area of crops	China statistical yearbook	The sown or transplanted area of crops on all land that should be harvested by agricultural producers and operators	10^3 ha	
Drought-affected area rate	—	The percentage of drought-affected area to total sown area of crops	%	
Drought-suffering area rate	—	The percentage of drought-suffering area to total sown area of crops	%	
Regional GDP	China Statistical Yearbook	The final result of production activities of all resident units in a region during a year calculated according to the market price	10^8 yuan	1979–2019 (Provincial scale)
Total population	China Statistical Yearbook	The population on December 31 of each year	10^4	1979–2019 (Provincial scale)
Rural population	China Statistical Yearbook	Total population except urban population	10^4	1979–2019 (Provincial scale)
Total power of agricultural machinery	China Statistical Yearbook	The sum of rated power of all agricultural machinery	10^3 kW	1979–2019 (Provincial scale)
Net amount of agricultural chemical fertilizer application	China Statistical Yearbook	The amount of chemical fertilizer actually used for agricultural production in a year	10^4 t	1979–2019 (Provincial scale)
Rural electricity use	China Statistical Yearbook	—	10^8 kW h	1979–2019 (Provincial scale)
Number of reservoirs	China Statistical Yearbook	—	—	1987–2019 (Provincial scale)
Reservoir storage capacity	China Statistical Yearbook	Total storage volume below check flood level	10^8 m^3	1987–2019 (Provincial scale)
Number of rural hydropower stations	China Statistical Yearbook	—	—	2011–2019 (Provincial scale)
Effective irrigation area	China Statistical Yearbook	The cultivated land area with certain water source, relatively flat land, supporting irrigation projects or equipment, and capable of irrigation in normal years	10^3 ha	1979–2019 (Provincial scale)
Number of ordinary high school graduates	China Statistical Yearbook	Indicating the level of education in a region	10^4	1987–2019 (Provincial scale)
Domestic patent application authorization	China Statistical Yearbook	The scientific, technological, and design achievements with independent intellectual property rights, indicating the level of science and technology in a region	—	1987–2019 (Provincial scale)
Total book prints	China Statistical Yearbook	Indicating the level of cultural propagation in a region	10^8	1979–2019 (Provincial scale)
Total water resources	China Water Resources Bulletin	The total amount of surface and subsurface water produced by local precipitation	10^8 m^3	2011–2019 (Provincial scale)
Water production modulus	China Water Resources Bulletin	The ratio of total water resources to regional area in a year	10^4 m^3/m^2	2011–2019 (Provincial scale)
Total water use	China Water Resources Bulletin	The gross water taken by various water users, including water transmission loss	10^8 m^3	2004–2019 (Provincial scale)
Agricultural water use	China Water Resources Bulletin	Including farmland irrigation water, forest and fruit land irrigation water, grassland irrigation water, fish pond replenishment water, and livestock and poultry water	10^8 m^3	2004–2019 (Provincial scale)
Actual irrigation water use per mu of cultivated land	China Water Resources Bulletin	The average water use per mu in actual agricultural irrigation area	m^3	2011–2019 (Provincial scale)
Effective utilization coefficient of farmland irrigation water	China Water Resources Bulletin	The proportion of actual water demand for crop growth in irrigation water	—	2011–2019 (Provincial scale)

(Continued on following page)

TABLE 1 | (*Continued*) List of indicators; their data sources and brief description are provided

Indicator	Data source	Description	Unit	Time period
Air temperature	China Climate Bulletin, ERA5	Mean annual air temperature	˚C	1979–2019 (Provincial scale)
Precipitation	China Climate Bulletin, ERA5	Mean annual precipitation	mm	1979–2019 (Provincial scale)

the functional relationship between potential factors and agricultural drought disasters. Finally, Gini importance, permutation feature importance, and accumulated local effects (ALE) plot were applied to identify principal factors and their influence on agricultural drought disasters.

The remainder of the article is organized as follows: **Section 2** provides an overview of study area and data sources, **Section 3** presents the methods adopted in the study, **Section 4** presents the results, **Section 5** discusses the findings and outlook, and finally the article is concluded in **Section 6**.

2 STUDY AREA AND DATA SOURCES

The paper selects China (excluding Hong Kong Special Administrative Region, Macao Special Administrative Region, and Taiwan Province) and its 31 provinces as the study area, and collects the annual data of 24 indicators at national or provincial spatial scale (**Table 1**). For national spatial scale, there are three indicators including drought-affected area, drought-suffering area, and total sown area of crops, whose data source is China Statistical Yearbook, with time scale from 1950 to 2020. For provincial scale, all the 24 indicators are included, but the time scale of different indicators is different, with the longest from 1979 to 2019 and the shortest from 2011 to 2019. Among the 24 indicators, data of 16 indicators of drought and social economy such as drought-affected area, regional GDP, and total book prints number are from China Statistical Yearbook; data of six indicators of water resources such as total water resources, total water use, and actual irrigation water use per mu of cultivated land are from China Water Resources Bulletin; data of two indicators including precipitation and temperature are from China Climate Bulletin and ERA5 (Hersbach et al., 2019). Previous studies indicated that ERA5 data have good suitability to China, which could reasonably depict regional difference of hydrological cycle elements (Su et al., 2020; Zhang et al., 2021; Zhou et al., 2021). However, we found that there is still a certain bias between original ERA5 products and *in situ* observations. We therefore corrected temperature and precipitation of the original ERA5 data by using bias correction method. It is found that the corrected ERA5 products are consistent with observations (**Supplementary Figure S1**).

Some data are missing for the 4 indicators of drought-affected area, drought-suffering area, number of reservoirs, and reservoir storage capacity. Among them, data of drought-affected area and drought-suffering area from 1967 to 1969 at national spatial scale are missing, and there are varying degrees of missing data for each province in each year. To ensure authenticity and objectivity, the paper does not deal with the missing data in drought-affected area and drought-suffering area. Data of reservoir number and reservoir storage capacity in 1999 are missing, and the arithmetic average values of 1998 and 2000 are used for interpolation.

3 METHODS

3.1 Mann-Kendall Test

Among all the trend analysis methods for time series, Mann–Kendall test is a method recommended by the World Meteorological Organization and widely used around the world. As a nonparametric test method, Mann–Kendall test is an effective tool to diagnose the trend of a data sequence, which does not need data series to follow a specific distribution, and is not affected by sample values, distribution types, and a few outliers either. Mann–Kendall test has the advantages of relatively simple calculation, high degree of quantification, and wide detection range. It is widely used in the analysis of hydrological data series. The mathematical fundamentals of Mann–Kendall test are as follows:

For time series $x_1, x_2, \ldots, x_n$, dual number p is calculated by

$$p = \sum_{i=1}^{n-1} \sum_{j-i+1}^{n} \mathbf{sgn}\left(x_j - x_i\right) \tag{1}$$

where $\mathbf{sgn}(x_j - x_i) = \begin{Bmatrix} 1 & if & x_j - x_i > 0 \\ 0 & if & x_j - x_i \le 0 \end{Bmatrix}$.

Then, τ, Var (τ), and U are calculated based on p:

$$\tau = \frac{4p}{n(n-1)} - 1 \tag{2}$$

$$\mathbf{Var}(\tau) = 2(2\mathbf{n}+5)/9\boldsymbol{n}(\boldsymbol{n}+1) \tag{3}$$

$$U = \frac{\tau}{\mathbf{Var}(\tau)^{1/2}} \tag{4}$$

When the value of statistic U is positive, it indicates that the time series shows an upward trend, otherwise the time series shows a downward trend. Taking the significance level as 95%, $U_{0.05}$ equals ±1.96. If $|U| \ge 1.96$, then the trend of the time series is significant, otherwise the trend of the time series is not significant.

3.2 Random Forest Model

Random Forest (RF) is a statistical learning theory proposed by Breiman (2001). Compared with other machine learning algorithms, Random Forest is insensitive to multicollinearity, robust to missing data and unbalanced data, does not need data preprocessing, and can provide reasonable prediction results for nonlinear relationships. It is one of the best

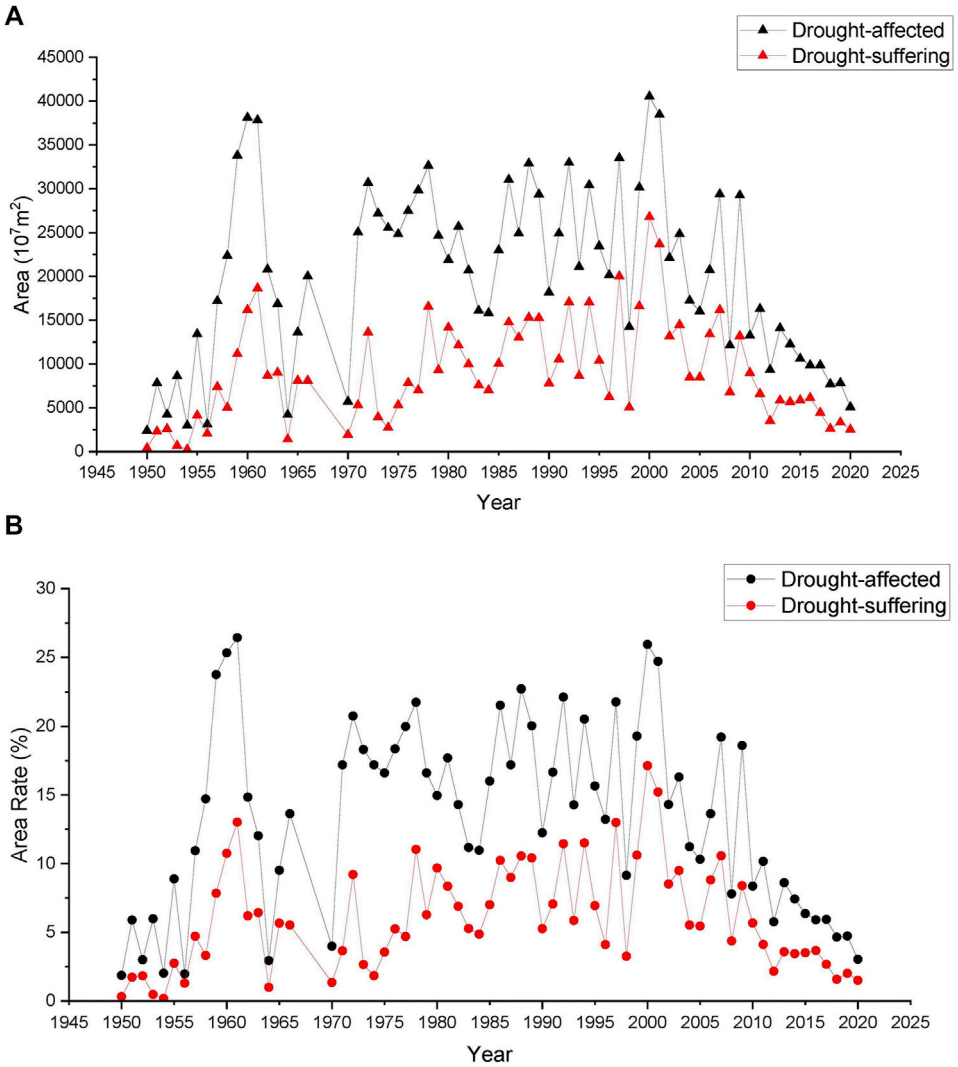


FIGURE 1 | Temporal variation of agricultural drought disasters in China from 1950 to 2020. **(A)** Drought-affected area and drought-suffering area. **(B)** Drought-affected area rate and drought-suffering area rate.

algorithms for processing high-dimensional data based on machine learning.

Random Forest is a collection of decision trees, each of which is slightly different from another. Random Forest repeatedly extracts samples from the training set, and the unselected samples constitute the out-of-bag data. Each training sample set is used to construct a decision tree. During the growth of a decision tree, features are randomly selected at each node, and the error of a decision tree is estimated according to the data out of bag. The prediction result of a RF algorithm is the mean value of the prediction result of each decision tree, and the prediction accuracy of a RF algorithm is estimated by the average prediction accuracy of each decision tree. The model construction process is as follows:

1) Split the whole data set into training set (75%) and test set (25%) randomly.
2) The training set is used to construct the RF model, and the best parameters of the model are determined based on network search and cross-validation. The accuracy of RF algorithm output mainly depends on three parameters: a) the number of trees (n_estimators) to grow in the forest, b) the maximum number of randomly selected features (max_features) at each node, and c) the maximum depth of each tree to grow (max_depth). In this paper, we randomly resampled different combinations of parameter sets with max_features ranging from one to total variables considered, and max_depth ranging from one to ten to avoid overfitting. Besides, we set n_estimators to 1,000 as suggested by Hengl et al. (2018) and Probst et al. (2017). Determination coefficient R^2 is adopted to measure the model training accuracy and then select the optimal parameters of the RF model. Calculation details are as follows:

TABLE 2 | Statistics of drought-affected area rate and drought-suffering area rate based on Mann–Kendall test in 1950–1980, 1981–2000, and 2001–2020

Stage	Item	Statistics			
		Mean (%)	SD	*Cv*	*U*
1950–1980	Drought-affected area rate	13.19	7.56	0.57	3.16[a]
	Drought-suffering area rate	4.72	3.56	0.75	2.61[a]
1981–2000	Drought-affected area rate	17.12	4.56	0.27	0.84
	Drought-suffering area rate	8.44	3.44	0.41	1.3
2001–2020	Drought-affected area rate	10.35	5.76	0.56	−4.48[a]
	Drought-suffering area rate	5.51	3.60	0.65	−4.54[a]

[a]*Means passing the significance test with a significance level of 5%.*

$$R^2 = \left[\frac{\mathrm{cov}(y_i, y_{pred})}{\sigma(y_i)\cdot\sigma(y_{pred})}\right]^2 \quad (5)$$

where y_i is the actual value in the validation set and y_{pred} is the predicted value of a RF model in the validation set. R^2 is between 0 and 1; the optimal value of R^2 is 1. Generally, R^2 greater than 0.6 would point to a model with good predictive power.

3) The test set is used to evaluate the generalization ability of the established RF model, and R^2, root mean square error (RMSE), and percent bias (PBIAS) are adopted to be the evaluation index. Calculation details are as follows:

$$\mathbf{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[y_i - y_{pred}\right]^2} \quad (6)$$

$$\mathbf{PBIAS} = 100 \times \frac{\sum_{i=1}^{n}(y_i - y_{pred})}{\sum_{i=1}^{n} y_i} \quad (7)$$

where n is the number of samples in the test set.

RMSE measures the average distance between predicted values and actual ones. The optimal value of RMSE is 0, with low-magnitude values indicating accurate model simulation.

PBIAS measures the average tendency of predicted values larger or smaller than their actual ones. The optimal value of PBIAS is 0, with low-magnitude values indicating accurate model simulation. Positive values represent underestimation bias, whereas negative values represent overestimation bias.

In this paper, Sklearn package in Python is used to realize the construction and verification of RF model.

3.3 Feature Importance

Two kinds of methods including Gini importance and permutation feature importance are adopted to filter out the significant features of target variables.

3.3.1 Gini Importance

Gini coefficient calculates the amount of probability of a specific feature that is classified incorrectly when selected randomly. Supposing there are K categories, the probability of i category is P_i, then the expression of Gini coefficient is

$$\mathbf{Gini}(P) = \sum_{i=1}^{K} P_i(1 - P_i) = 1 - \sum_{i=1}^{K} P_i^2 \quad (8)$$

Gini coefficient reflects the impurity at a node in a decision tree. Each time a particular feature is used to split a node, the Gini coefficient for the child nodes are calculated and compared with that of the original node. If the reduction is large, it shows that the feature at the node has a great impact on the decision tree. The importance of a feature is computed as the normalized total reduction of the Gini coefficient brought by that feature in a Random Forest, and the sum of Gini importance of all features is equal to 1. In this paper, Sklearn package in Python is used to realize the Gini feature importance.

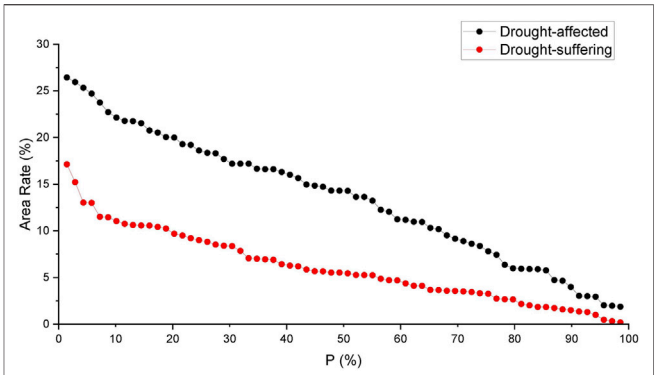


FIGURE 2 | Frequency curves of drought-affected area and drought-suffering area in China.

3.3.2 Permutation Feature Importance

Permutation feature importance, introduced by Breiman (2001) for Random Forests, measures the increase in the prediction error of the model after we permuted the feature's values. Features with higher importance value likely have higher dominant control. Permutation feature importance provides a highly compressed, global insight into the model's behavior. Detailed algorithm turns to Fisher et al. (2019) and Molnar (2021). In this paper, eli5 package in Python is used to realize permutation feature importance.

3.4 Accumulated Local Effects Plot

Accumulated local effects plot describes how features influence the prediction of RF model on average, which can present the nonlinear response of target variables to features. Compared with partial dependence plots (PDPs), ALE plots can still work when features are correlated. In addition, the value at each point of the ALE curve is the difference to the mean prediction, which makes their interpretation clearer. In this paper, we repeated the ALE algorithm in Monte Carlo simulation with 100 replicates, where on each replicate we generated a new training data set and refit the RF model with the same best parameters. Detailed algorithm turns to Grömping (2020) and Molnar (2021). In this paper, alepython package in Python is used to realize ALE plot.

4 RESULTS

4.1 Spatial-Temporal Pattern of Agricultural Drought Disasters

4.1.1 Temporal Evolution

The temporal variations of drought-affected area and drought-suffering area, drought-affected area rate and drought-suffering area rate over the years from 1950 to 2020 are shown in **Figure 1**.

TABLE 3 | Frequency characteristic value of drought-affected area and drought-suffering area in China

Possibility	75%	50%	25%	10%	5%
Drought-affected area rate	8.22	14.29	18.42	21.88	24.37
Drought-suffering area rate	3.29	5.49	8.86	10.83	12.47

It can be seen that drought-affected area and drought-suffering area, drought-affected area rate and drought-suffering area rate fluctuate since 1950, which can be roughly divided into three stages: 1950–1980, 1981–2000, and 2001–2020. By calculating the statistics of the time sequences of drought-affected area rate and drought-suffering area rate (**Table 2**), it can be found that the national multi-year average drought-affected area rate and drought-suffering area rate are 13.19 and 4.72% in 1950–1980, 17.12 and 8.44% in 1981–2000, and 10.35 and 5.51% in 2001–2020. Compared among the three stages, the national multi-year average drought-affected area rate and drought-suffering area rate are the largest in 1981–2000, which are about 1.3 times and 1.8 times as large as those in 1950–1980, and 1.7 times and 1.5 times as large as those in 2001–2020. The variation coefficients of drought-affected area rate and drought-suffering area rate are 0.57 and 0.75 in 1950–1980, 0.27 and 0.41 in 1981–2000, and 0.56 and 0.65 in 2001–2020. Compared among the three stages, the variation coefficients of drought-affected area rate and drought-suffering area rate are the smallest in 1981–2000. This indicates that the agricultural drought disasters in 1950–1980 and 2001–2020 are relatively gentle, but the interannual variation is relatively large; the agricultural drought disasters in 1981–2000 are relatively heavy, but the interannual variation is relatively small. The U statistics derived from the Mann–Kendall test of drought-affected area rate and drought-suffering area rate are 3.16 and 2.61 in 1950–1980 with significant upward trend, 0.84 and 1.3 in 1981–2000 with non-significant upward trend, and −4.48 and −4.54 in 2001–2020 with significant downward trend. Compared among the three stages, the downward trend is the most obvious in 2001–2020. In the past decade (2011–2020), drought-affected area rate and drought-suffering area rate have been decreasing continuously, with an average drought-affected area rate of 6.26% and an average drought-suffering area rate of 2.82%.

The frequency curves of drought-affected area rate and drought-suffering area rate are shown in **Figure 2**, and the frequency characteristic values are shown in **Table 3**. The variation of frequency curves of drought-affected area rate and drought-suffering area rate is overall gentle, but the variation of rare drought ($p < 10\%$) is relatively heavy. For drought-affected area rate, there is a probability of 50% greater than 14.29%, 10% greater than 21.88%, and 5% greater than 24.37%. For drought-suffering area rate, there is a probability of 50% greater than 5.49%, 10% greater than 10.83%, and 5% greater than 12.47%. Under the probability of 50%, the value of drought-affected area rate is about 2.6 times as large as that of drought-suffering area rate, while under the probability of 10%, the value of drought-affected area rate is only about 2.0 times as large as that of drought-suffering area rate. This indicates that with the increase of drought intensity, it is more likely to evolve into a drought disaster of high impact with crop yield 30% or more lower than normal year. The years 1961, 2000, 1960, 2001, and 1959 are the 5 years with the highest drought-affected area rate in China from 1950 to 2020. Also, 2000, 2001, 1961, 1997, and 1994 are the 5 years with the highest drought-suffering area rate in China from 1950 to 2020. The reduction of grain yield in these years was over 5%, up to 11.5% (Zhang et al., 2008).

4.1.2 Spatial Pattern

The spatial distribution of the average drought-affected area rate and drought-suffering area rate from 1979 to 2019 in China is shown in **Figure 3**, and we can see that on the whole, drought-affected area rate and drought-suffering area rate in northern China are higher than those in southern China. In southern China, the southeast coastal provinces have the lowest drought-affected area rate and drought-suffering area rate. In Northern China, drought-affected area rate and drought-suffering area rate of Beijing, Tianjin, and Xinjiang are relatively low. The top five

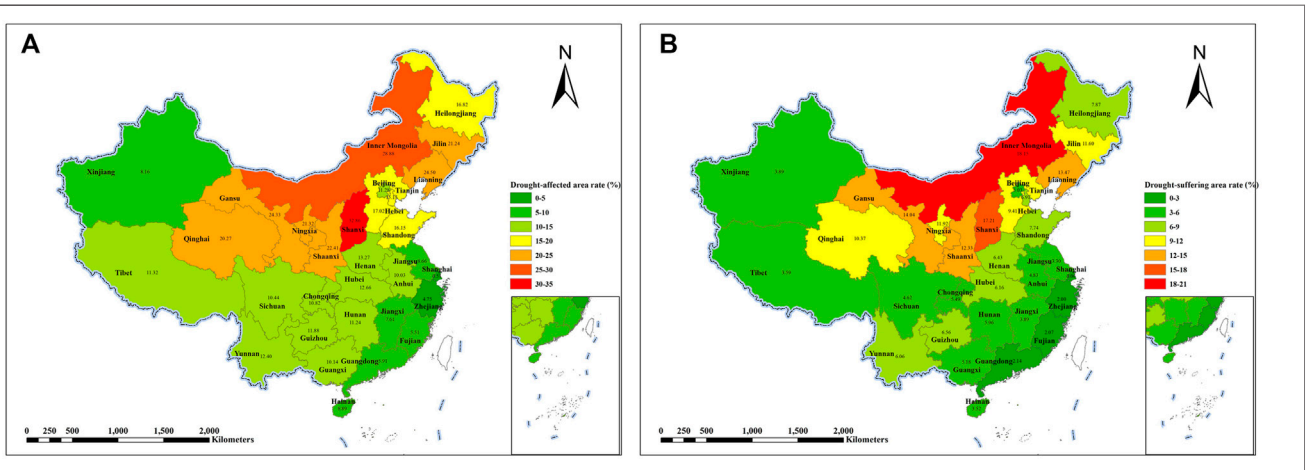


FIGURE 3 | Spatial distribution of multi-year average agricultural drought disasters. **(A)** Drought-affected area rate. **(B)** Drought-suffering area rate.

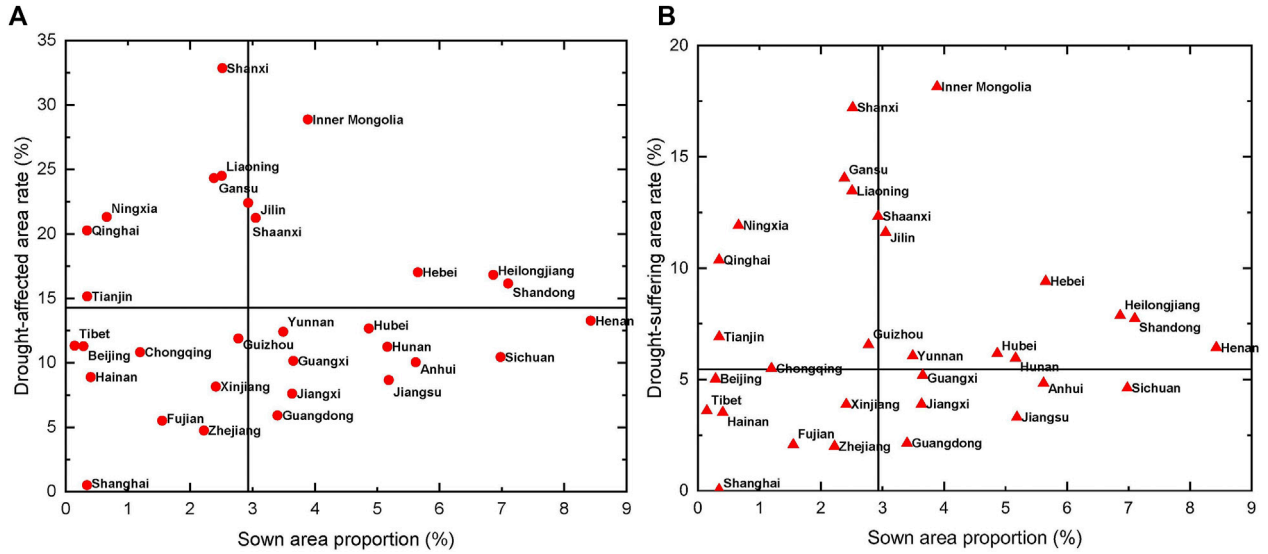


FIGURE 4 | Agricultural drought disasters against sown area proportion. **(A)** Drought-affected area rate. **(B)** Drought-suffering area rate.

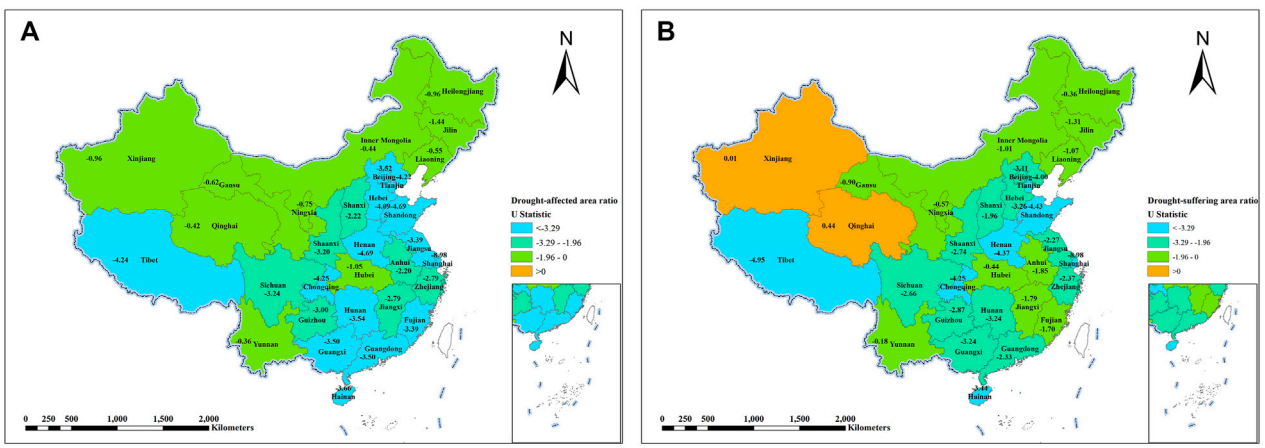


FIGURE 5 | Spatial distribution of *U* statistic derived from Mann–Kendall test of agricultural drought disasters. **(A)** Drought-affected area rate. **(B)** Drought-suffering area rate.

provinces with the highest drought-affected area rate are Shanxi (32.86%), Inner Mongolia (28.88%), Liaoning (24.5%), Gansu (24.33%), and Shaanxi (22.41%). The top five provinces with the highest drought-suffering area rate are Inner Mongolia (18.15%), Shanxi (17.21%), Gansu (14.04%), Liaoning (13.47%), and Shaanxi (12.33%). The two collections of provinces are exactly the same and all of them are located in northern China.

Scatter plots of drought-affected area rate and drought-suffering area rate against sown area proportion are shown in **Figure 4**. Here, sown area proportion refers to the percentage of sown area of crops in a province in the total sown area of crops in China, which can reflect the contribution of agricultural production of a province to the whole country. The lines represent the 50% quantile of drought-affected area rate, drought-suffering area rate, and sown area proportion, which is 14.29, 5.49, and 2.93%, respectively. Among the four quadrants, the first quadrant indicates that agricultural drought disasters are relatively serious and sown area of crops accounts for a high proportion of the country. The second quadrant indicates that agricultural drought disasters are relatively serious but sown area of crops accounts for a low proportion of the country. The third quadrant indicates that sown area of crops accounts for a low proportion of the country and agricultural drought disasters are relatively light. The fourth quadrant indicates that sown area of crops accounts for a high proportion of the country but agricultural drought disasters are relatively light. Therefore, the provinces located in the first quadrant need to be paid enough attention. As we can see from the figure, seven provinces including Inner Mongolia, Gansu, Jilin, Shaanxi, Hebei, Heilongjiang, and Shandong account for a high proportion of

TABLE 4 | Index system of influencing factors of agricultural drought disasters

Index system				Data series			
Abbr	**Primary**	**Secondary**	**Tertiary**	**1979–2019**	**1987–2019**	**2004–2019**	**2011–2019**
F1	Natural factor	Meteorological conditions	Air temperature	1	1	1	1
F2			Precipitation	1	1	1	1
F3		Water resources conditions	Total water resources	0	0	0	1
F4			Water production modulus	0	0	0	1
F5	Human factor	Basic socio-economic conditions	Regional GDP	1	1	1	1
F6			Total population	1	1	1	1
F7			Total water use	0	0	1	1
F8		Agricultural development	Rural population	1	1	1	1
F9			Total sown area of crops	1	1	1	1
F10			Total power of agricultural machinery	1	1	1	1
F11			Net amount of agricultural chemical fertilizer application	1	1	1	1
F12			Agricultural water use	0	0	1	1
F13			Rural electricity use	1	1	1	1
F14		Water conservancy conditions	Number of reservoirs	0	1	1	1
F15			Reservoir storage capacity	0	1	1	1
F16			Number of rural hydropower stations	0	0	0	1
F17			Effective irrigation area	1	1	1	1
F18			Actual irrigation water use per mu of cultivated land	0	0	0	1
F19			Effective utilization coefficient of farmland irrigation water	0	0	0	1
F20		Water saving consciousness	Number of ordinary high school graduates	0	1	1	1
F21			Domestic patent application authorization	0	1	1	1
F22			Total book prints	1	1	1	1
Total	—	—	—	11	15	17	22

sown area and high drought-affected area rate. Ten provinces including Inner Mongolia, Jilin, Shaanxi, Hebei, Heilongjiang, Shandong, Yunnan, Hubei, Hunan, and Henan account for a high proportion of sown area and high drought-suffering area rate.

The trend of drought-affected area rate and drought-suffering area rate with time of each province was estimated based on Mann–Kendall test (**Figure 5**). The drought-affected area rate of all provinces presents a downward trend with time. Heilongjiang, Jilin, Liaoning, Inner Mongolia, Ningxia, Gansu, Qinghai, Xinjiang, Hubei, and Yunnan show a non-significant downward trend, mainly distributed in northeast China and northwest China. The rest of the provinces show a significant downward trend. The drought-suffering area rate of Xinjiang and Qinghai presents a non-significant upward trend with time, while the drought-suffering area rate of the other 29 provinces presents a downward trend with time. Heilongjiang, Jilin, Liaoning, Inner Mongolia, Ningxia, Gansu, Hubei, Anhui, Jiangxi, Fujian, and Yunnan show a non-significant downward trend, mainly distributed in northeast China, northwest China, and provinces in the middle and lower reaches of the Yangtze River. The rest of the provinces show a significant downward trend.

4.2 Analysis on the Influencing Factors of Agricultural Drought Disasters

4.2.1 Index System

Agriculture drought disasters, impacted by many natural factors and human activities, are a complicated system. To explore the principal influencing factors and their effect on agriculture drought disasters, it is necessary to select operational indexes comprehensively and construct scientific multi-hierarchy index systems. Generally, agricultural drought disasters are affected by natural and human factors.

Natural factors are mainly divided into meteorological conditions and water resources conditions, including temperature, precipitation, total water resources, and water production modulus, which reflects the basic hydrothermal conditions of agricultural production in a region.

Human factors are mainly divided into 4 aspects: basic socio-economic conditions, agricultural development, water conservancy conditions, and water saving consciousness. Basic socio-economic conditions include three indexes, regional GDP, total population, and total water use, which represent the comprehensive socio-economic capacity in a region. Agricultural development includes six indexes, rural population, total sown area of crops, total power of agricultural machinery, net amount of agricultural chemical fertilizer application, agricultural water use, and rural electricity use, which shows the basic conditions of agricultural production in a region. Water conservancy conditions include six indexes, number of reservoirs, reservoir storage capacity, number of rural hydropower stations, effective irrigation area, actual irrigation water use per mu of cultivated land, and effective utilization coefficient of farmland irrigation water, which indicates the development and utilization level of water resources and the ability of drought prevention and reduction in a region. Water saving consciousness includes three indexes,

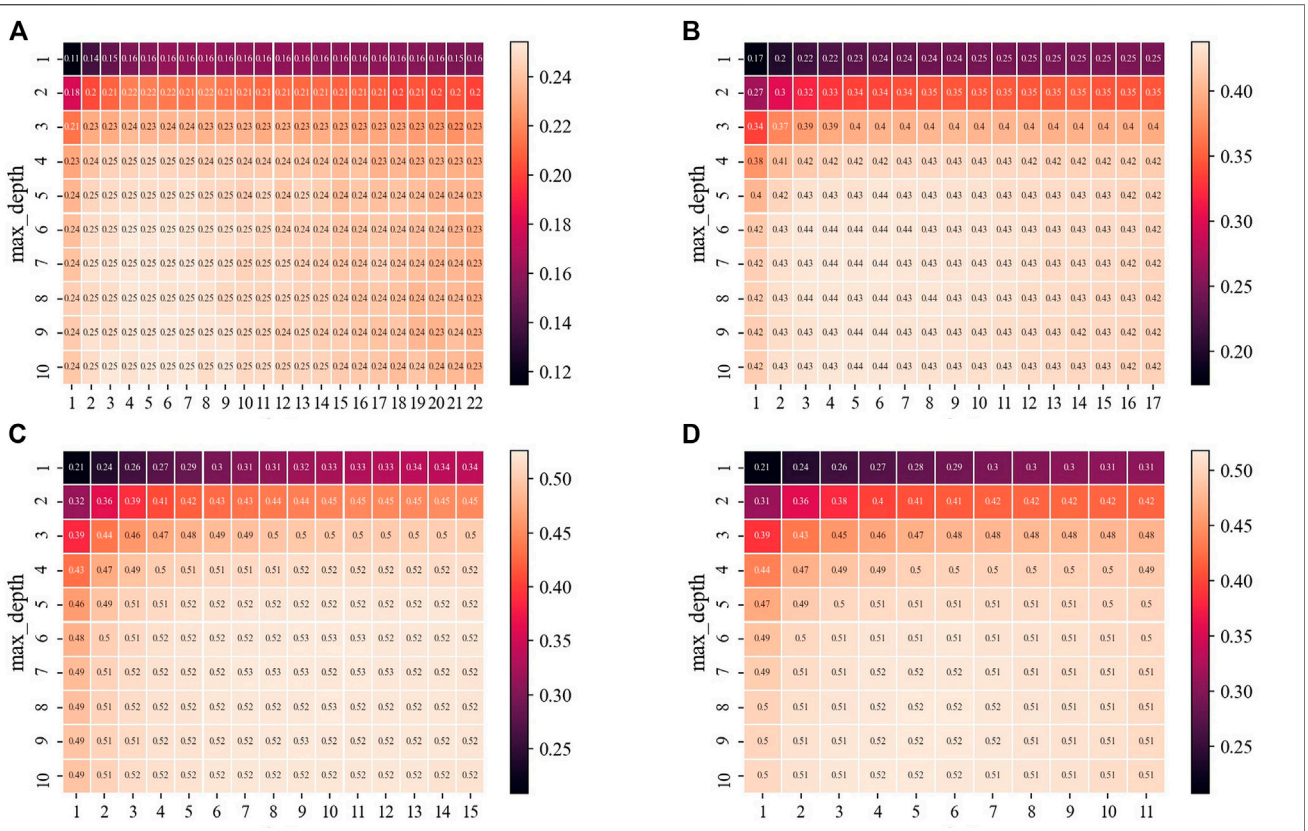


FIGURE 6 | Heatmaps are provided with R^2 of grid research and cross-validation in case of drought-affected area. **(A)** 2011–2019. **(B)** 2004–2019. **(C)** 1987–2019. **(D)** 1979–2019.

TABLE 5 | Optimal parameters and the corresponding model performance metrics of RF algorithms for drought-affected area and drought-suffering area

Drought characteristics	Time series	Best R^2	Optimal RF parameters		Model performance		
			max_depth	max_features	R^2	RMSE	PBIAS
Drought-affected area	2011–2019	0.25	10	5	0.26	1.30	−4.40
	2004–2019	0.44	7	6	0.38	1.24	5.84
	1987–2019	0.53	7	9	0.60	1.01	−0.44
	1979–2019	0.52	8	5	0.65	0.88	−0.04
Drought-suffering area	2011–2019	0.26	5	3	0.21	1.21	−4.30
	2004–2019	0.43	7	4	0.42	1.22	7.88
	1987–2019	0.50	7	10	0.59	1.07	−0.49
	1979–2019	0.47	9	5	0.62	0.96	0.27

number of ordinary high school graduates, domestic patent application authorization, and total book prints, which indirectly reflect the level of education, scientific research, and cultural propagation in a region, respectively.

Thus, an index system of influencing factors of agricultural drought disasters with 22 indexes was established (**Table 4**). Due to the limited time length of some indexes obtained from data sources (**Table 1**), a total of four datasets have been integrated. Dataset of 2011–2019 has the shortest time length (9 years), but contains all the 22 indexes. Dataset of 2004–2019 has time length of 16 years with 17 indexes. Dataset of 1987–2019 has time length of 33 years with 15 indexes. Dataset of 1979–2019 has the longest time length (41 years), but contains only 11 indexes. In the following paper, we will use these four datasets to explore the influence of sample number and feature number on the fit goodness of Random Forest algorithm.

4.2.2 RF Model Performance

Taking drought-affected area and drought-suffering area as target variables, and taking indexes in **Table 4** as features, Random Forest model was constructed based on the method described in **Section 3.4**. It should be mentioned that log (y+1) transformation

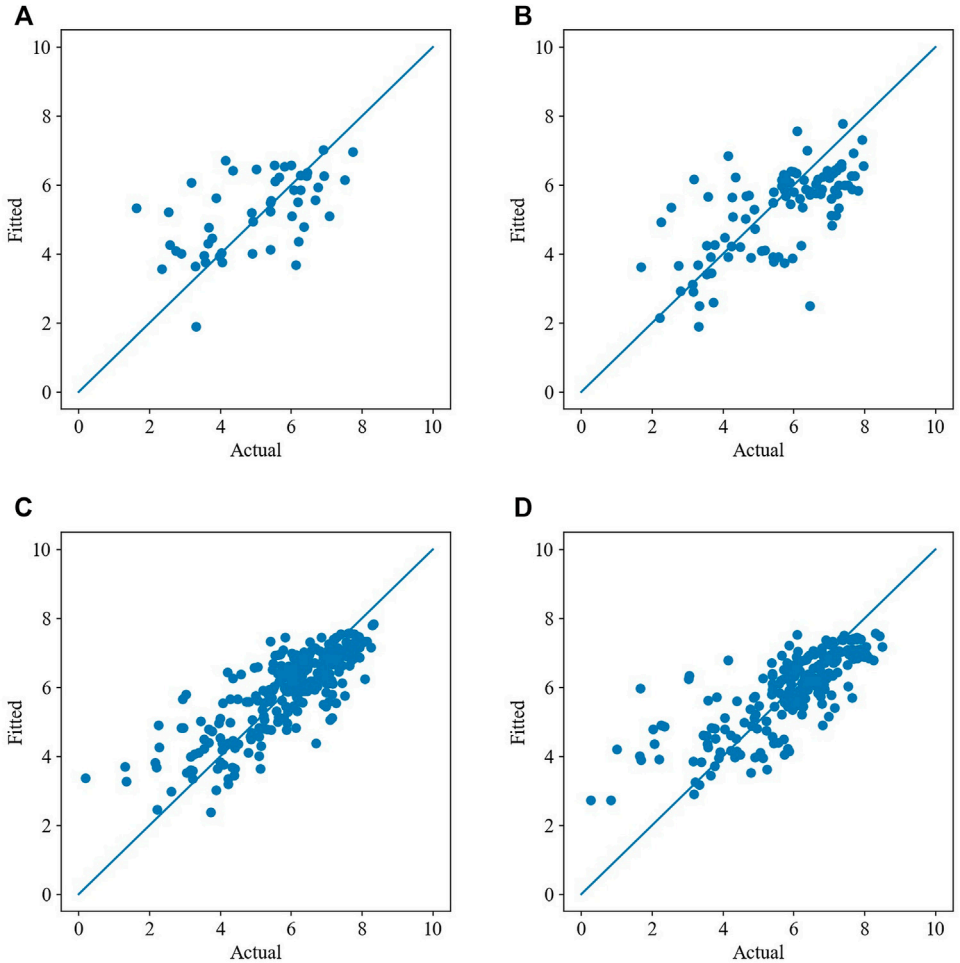


FIGURE 7 | Comparison plots are provided with *x*-axis as the actual values in test set of drought-affected area and *y*-axis as the fitted values of the RF model. **(A)** 2011–2019. **(B)** 2004–2019. **(C)** 1987–2019. **(D)** 1979–2019.

was carried out on target variables when building the RF model because both drought-affected area and drought-suffering area correspond to a positive skew distribution (**Figure 2**).

To explore the influence of sample number and feature number on the fit goodness of RF algorithm, the four datasets of 2011–2019, 2004–2019, 1987–2019, and 1979–2019 were used as the input data, respectively. **Figure 6** and **Supplementary Figure S2** present the results of parameter network search and cross-validation for drought-affected area and drought-suffering area in the training set. It can be found that when max_depth is small (1–4), R^2 derived from all the four datasets is poor. When max_features is small (1–2), R^2 derived from all the four datasets is poor. For the dataset of 2011–2019, R^2 is poor when max_features is large (>7). For the datasets of 2004–2019, 1987–2019, and 1979–2019, R^2 changes little with max_depth (>4) and max_features (>2) increasing.

According to the best R^2 of cross-validation, we obtain the optimal parameters of RF algorithms derived from the four datasets (**Table 5**). It shows that the optimal max_depth of all the algorithms is relatively large (5–10), and the optimal max_features of the algorithms varies greatly (from 3 to 10). For drought-affected area, the best R^2 of dataset of 2011–2019 is the smallest (0.25), and that of 1979–2019 is the largest (0.52). For drought-suffering area, the best R^2 of dataset of 2011–2019 is the smallest (0.26), and that of 1987–2019 is the largest (0.50). Generally speaking, the increase of sample numbers significantly improves the simulation accuracy of the RF algorithm.

We evaluated the performance of RF algorithm derived from the four datasets in the test set. The metrics of R^2, RMSE, and PBIAS for the corresponding model configurations are listed in **Table 5**. Among the four datasets, the algorithm derived from the dataset of 1979–2019 performs best with the largest R^2, smallest RMSE and PBIAS closest to 0, indicating that the RF algorithm driven by the dataset of 1979–2019 can explain the variance of drought-affected area and drought-suffering area to the greatest extent. It can also be seen from **Figure 7D** and **Supplementary Figure S3D** that the deviation between the actual values in the test set and the fitted values of the RF algorithm are the smallest. The RF algorithm driven by the dataset of 2011–2019 performs

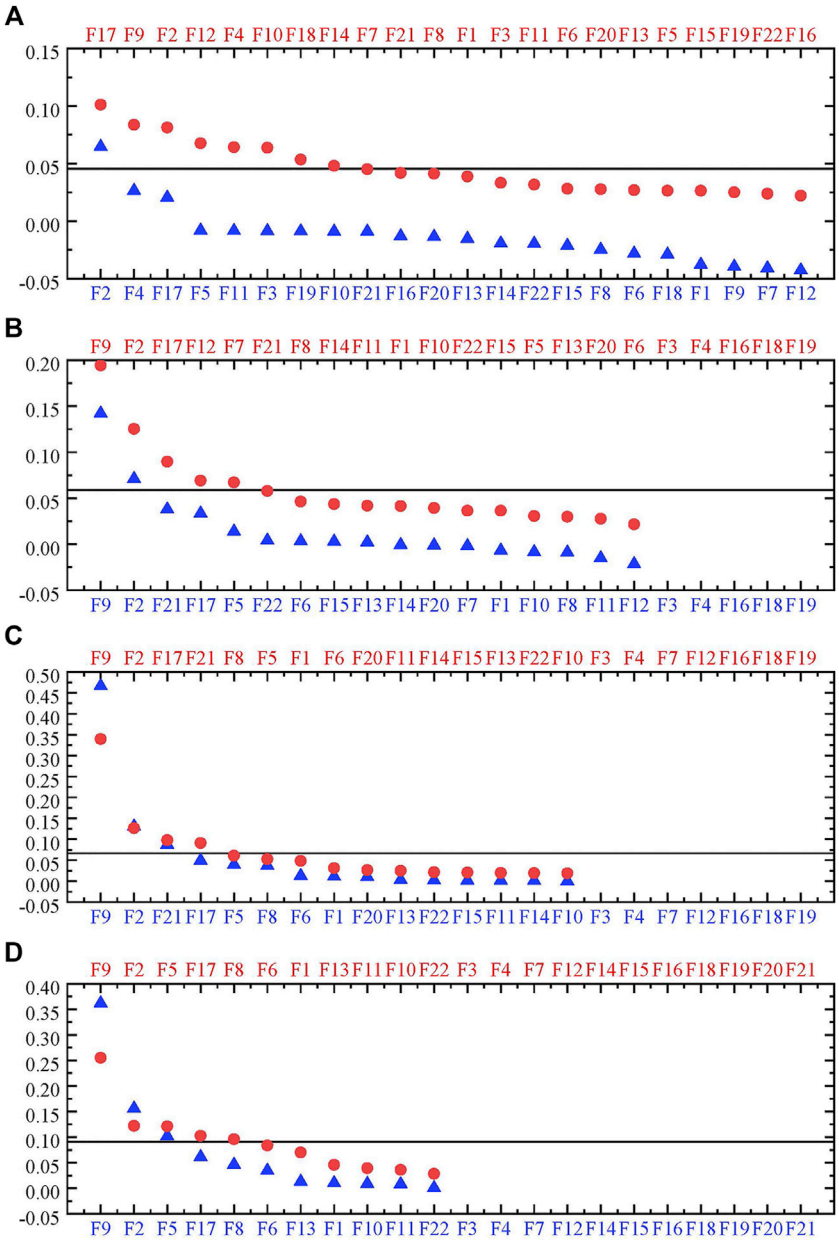


FIGURE 8 | Rank plots are provided in descending order with features exhibiting maximum importance on the leftmost in case of drought-affected area based on datasets of **(A)** 2011–2019, **(B)** 2004–2019, **(C)** 1987–2019, and **(D)** 1979–2019. Red circles and top *x*-axis represent results derived from Gini importance, blue triangles and bottom *x*-axis represent results derived from permutation feature importance. The line shows the average feature importance, which can be used as a threshold to determine the significant features of interest for Gini importance.

poorly with R^2 smaller than 0.30 and overestimates the actual values (**Figure 7A** and **Supplementary Figure S3A**). While the RF algorithm driven by the dataset of 2004–2019 underestimates the actual values (**Figure 7B** and **Supplementary Figure S3B**) with R^2 smaller than 0.45. The RF algorithm derived from the dataset of 1987–2019 has high prediction accuracy, but slightly overestimates the actual values (**Figure 7C** and **Supplementary Figure S3C**), especially drought with low intensity. The test result of the model also shows that the increase of sample numbers can improve the generalization ability of the RF algorithm compared with the increase of the number of features.

4.2.3 Drought-Affected Area

Figure 8 shows the rank of natural and human factors that has potential influence on drought-affected area indicated by the two methods of Gini importance and permutation feature importance based on the four datasets. Red circles and top *x*-axis represent results derived from Gini importance, and blue triangles and bottom *x*-axis represent results

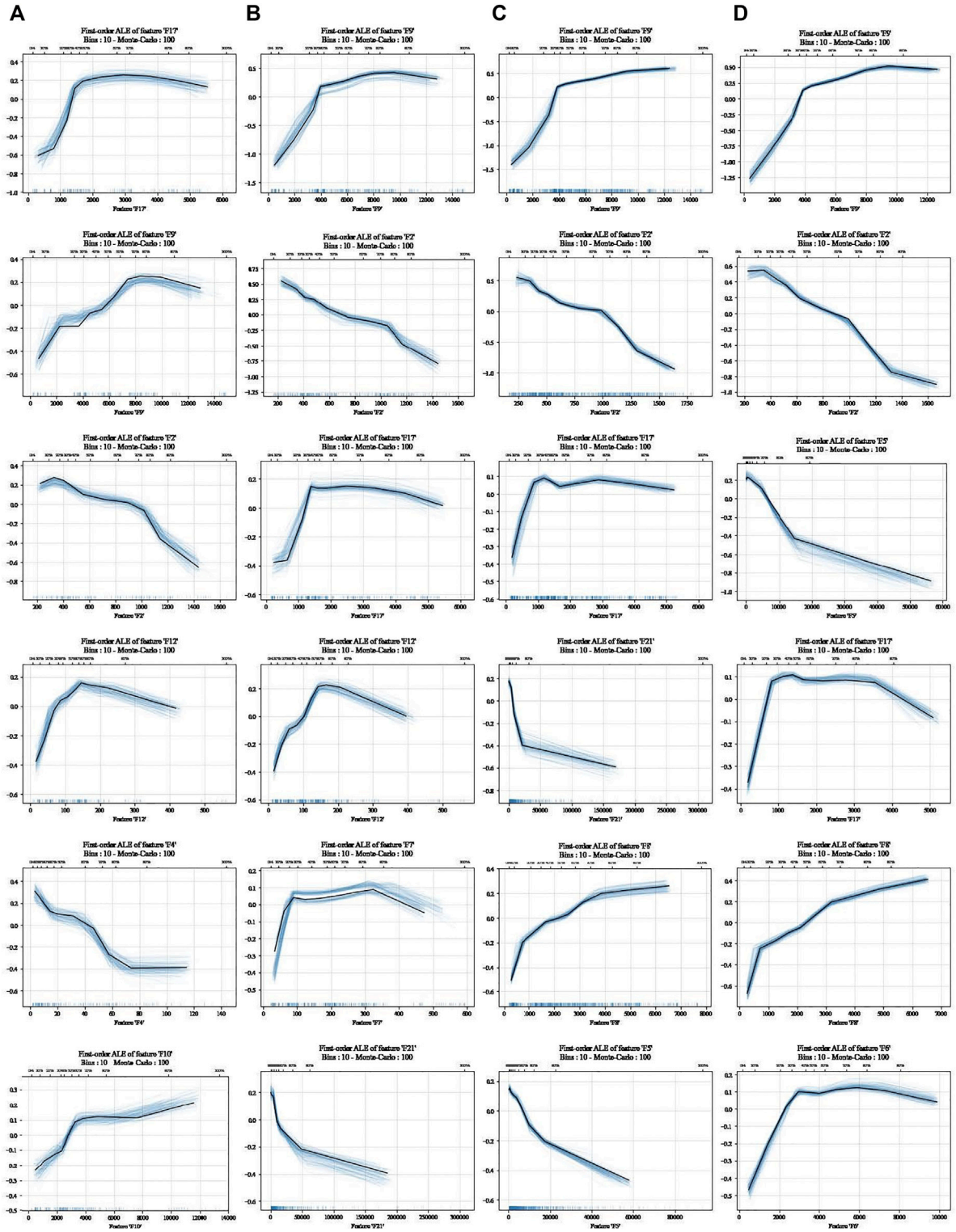


FIGURE 9 | ALE plots of top six features controlling drought-affected area. **(A)** 2011–2019. **(B)** 2004–2019. **(C)** 1987–2019. **(D)** 1979–2019.

derived from permutation feature importance. The line shows the average feature importance of Gini importance, which can be used as a threshold to determine the significant features of interest.

For the dataset of 2011–2019, F17 (effective irrigation area), F9 (total sown area of crops), F2 (precipitation), F12 (agricultural water use), F4 (water production modulus), F10 (total power of

agricultural machinery), F18 (actual irrigation water use per mu of cultivated land), and F14 (number of reservoirs) have more than average Gini importance, while F2, F4, and F17 show the top three permutation importance. It is worth noting that these three features are the only three features whose permutation importance value is over 0. For the dataset of 2004–2019, F9, F2, F17, F12, and F7 (total water use) have more than average Gini importance, while F9, F2, and F21 (domestic patent application authorization), and F17 and F5 (regional GDP) show the top five permutation importance. For the dataset of 1987–2019, F9, F2, F21, and F17 have more than average Gini importance; these four features are exactly the same as the top four features derived from permutation feature importance. For the dataset of 1979–2019, F9, F2, F5, F17, and F8 (rural population) have more than average Gini importance; these five features are exactly the same as the top five features derived from permutation feature importance.

We can find that the significant features screened by the two methods of Gini importance and permutation importance based on the datasets of 1979–2019 and 1987–2019 are identical, while the significant features screened by the two methods derived from the dataset of 2011–2019 show significant difference. Considering the performance of RF models, we can deduce that models with low prediction accuracy have high uncertainties, and the feature importance derived from the datasets of 1979–2019, 1987–2019, and 2004–2019 is more reliable.

In addition, F9, F2, and F17 dominate in controlling drought-affected area for all the four datasets. F21 dominates in controlling drought-affected area for datasets of 1987–2019 and 2004–2019. F5 dominates in controlling drought-affected area for datasets of 1979–2019 and 2004–2019. Based on the aforementioned facts, we can make a preliminary judgement that F9, F2, F17, F21, and F5 are the principal features influencing drought-affected area.

To further understand how these key features influence drought-affected area, the ALE plots of the top six features derived from the four datasets are plotted (**Figure 9**). Basically, drought-affected area is positively correlated with F9, and the nonlinear relation between F9 and drought-affected area can be divided into two stages at 4 million hectares. Before 4 million hectares, drought-affected area increases significantly with the increase of F9; after 4 million hectares, drought-affected area increases slightly or even decreases with the increase of F9. Drought-affected area is negatively correlated with F2 based on a linear behavior. A non-significant jump point can be observed at 1,000 mm, and the downward trend after 1,000 mm is steeper than that before 1,000 mm. The relation between drought-affected area and F17 is complicated, which can be roughly divided into two stages around 1 million hectares. Before 1 million hectares, drought-affected area increases significantly with the increase of F17; after 1 million hectares, drought-affected area decreases slightly with the increase of F17. Two-stage relationship can be observed between drought-affected area and F21. Before 20,000, drought-affected area decreases significantly over F21 increasing; after 20,000, drought-affected area decreases slightly over F21 increasing. Drought-affected area is negatively correlated with F5. A jump point can be observed around 150 trillion yuan, and the downward trend after 150 trillion yuan is gentler than that before 150 trillion yuan.

4.2.4 Drought-Suffering Area

Supplementary Figure S4 shows the rank of natural and human factors that has potential influence on drought-suffering area indicated by the two methods of Gini importance and permutation feature importance based on the four datasets. For the dataset of 2011–2019, F17, F9, F12, F2, F4, F10, F7, F1 (air temperature), and F14 have more than average Gini importance, while F4, F2, F17, and F3 (total water resources) show the top four permutation importance. It should be noted that these four features are the only four features whose permutation importance value is over 0. For the dataset of 2004–2019, F9, F2, F17, F7, F12, and F21 have more than average Gini importance, while F9, F21, F2, F17, F5, and F6 (total population) show the top six permutation importance. For the dataset of 1987–2019, F9, F2, F21, and F17 have more than average Gini importance; these four features are exactly the same as the top four features derived from permutation feature importance. For the dataset of 1979–2019, F9, F2, F5, and F17 have more than average Gini importance; these four features are exactly the same as the top four features derived from permutation feature importance.

In addition, F9, F2, and F17 dominate in controlling drought-suffering area for all the four datasets. F21 dominates in controlling drought-suffering area for datasets of 1987–2019 and 2004–2019. F5 dominates in controlling drought-suffering area for datasets of 1979–2019 and 2004–2019. Based on the aforementioned facts, we can make a preliminary judgement that F9, F2, F17, F21, and F5 are the key features influencing drought-suffering area. To further understand how these dominant features influence drought-suffering area, the ALE plots of top six features for the four datasets are plotted (**Supplementary Figure S5**). The nonlinear relation between drought-suffering area and F9, F2, F17, F21, and F5 is almost the same as drought-affected area, so we will not go into much detail here.

5 DISCUSSION

Over the past 70 years, agricultural drought disasters in China have undergone significant fluctuations, and the spatial pattern of agricultural drought disasters varies from north to south. In terms of drought intensity characterized by the absolute value of drought-affected area and drought-suffering area, northern China is significantly higher than southern China. In terms of the impact of drought on agricultural production characterized by the relationship between drought-affected/suffering area rate against sown area proportion, eight northern provinces including Inner Mongolia, Gansu, Jilin, Shaanxi, Hebei, Heilongjiang, Shandong, and Henan and three southern provinces including Yunnan, Hubei, and Hunan are greatly affected by drought. In terms of the trend of drought characterized by the U statistics derived from Mann–Kendall test, among the aforementioned 11 provinces, drought in Inner Mongolia, Gansu, Jilin, Heilongjiang, Yunnan, and Hubei shows a non-significant downward trend,

while drought in Shaanxi, Hebei, Shandong, Hunan, and Henan presents a significant downward trend. These results show that we should carry out regular monitoring, predicting, early warning, and risk management on drought in northern China. At the same time, we should also pay enough attention to the three southern provinces including Yunnan, Hubei, and Hunan. In addition, it should be noted that with the increase of drought intensity, it is more likely to evolve into a drought disaster of high impact with crop yield 30% or much lower than normal year, which indicates that all provinces should further improve their response ability to heavy drought disasters.

Wang et al. (2021a, 2021b) showed that standardized precipitation evapotranspiration index (SPEI) has the best correlation with drought-affected area in Northeast China, and by analyzing the trend of SPEI during the growing season, they found that there was an overall increasing trend, and the jump point was in 1983. Liao and Zhang (2017) found that from the perspective of drought-affected area rate, northern China is relatively high, while the middle and lower reaches of the Yangtze River are relatively low, and the spatial distribution of drought-affected area rate is basically consistent with that of meteorological drought days. These results show that sometimes drought indices derived from hydro-meteorological variables can be a good representation for agricultural drought disaster. In fact, drought-affected area and drought-suffering area are not only related to drought itself, but also closely related to the local agricultural distribution characteristics.

By analyzing the feature importance in each dataset based on the two methods of Gini importance and permutation feature importance, we can deduce that total sown area of crops, precipitation, effective irrigation area, domestic patent application authorization, and regional GDP are the top five dominant factors that influence agricultural drought disasters. These 5 factors correspond to 5 secondary indexes including agriculture development, meteorological conditions, water conservancy conditions, water saving consciousness, and basic socio-economic conditions, respectively. This also illustrates the complexity of the influencing factors of agricultural drought disasters and indicates the applicability of the index system constructed in this paper.

Qian et al. (2016) analyzed the relationships between agricultural drought and climate factors based on Pearson correlation coefficient using vegetation condition index data recorded from 1982 to 2010; they found that temperature and wind velocity were the main factors that influenced drought in the agricultural areas of China. Liu et al. (2021) analyzed meteorological factors on droughts in Xilinguole Grassland with a combination of Pearson correlation analysis and *t*-tests, and result shows that temperature, precipitation, water vapor pressure, and solar radiation are the key factors. These two studies only considered the effect of meteorological factors based on linear methods, and results showed that temperature and precipitation are the main influencing factors on agricultural drought, which is basically consistent with the results in our paper (**Figure 8** and **Supplementary Figure S4**). However, they neglected the effect of human activities on agricultural drought, as Qian et al. (2016) discussed that the correlation between VCI and precipitation was low, possibly due to the widespread use of artificial irrigation technology, which reduces the reliance of agricultural areas on precipitation. Zhang et al. (2021) identified influencing factors of regional agricultural drought vulnerability during the period from 2012 to 2018 in Henan Province based on grey trend relational analysis method, and they found that the influencing factors mainly related to natural resources, agricultural industrial structure, agricultural attention, agricultural water efficiency, residents' awareness of water conservation, and water and soil conservation measures. This study explored influencing factors of agricultural drought considering both natural and human factors based on a nonlinear method, which is quite similar to our research, and the important impact of natural resources, agricultural attention, agricultural water efficiency, and residents' awareness of water conservation is also indicated in our results. All aforementioned studies explored the dominant influencing factors on agricultural drought, and the results are different since influencing factors considered are different. However, few research studies have investigated how these factors impact agricultural drought.

ALE plots were adopted to interpret RF algorithm for revealing how key factors impact agricultural drought disasters in our paper, and we find that there is a negative linear correlation between agricultural drought disasters and precipitation basically, but a non-significant jump point can still be observed at 1,000 mm, showing that when precipitation exceeds 1,000 mm, the intensity of agricultural drought disasters will decrease more significantly. This is an important reason for the low intensity of agricultural drought disasters in southeast coastal China, such as Jiangsu, Shanghai, Zhejiang, Fujian, Guangdong, and Hainan. Moreover, agricultural drought disaster negatively correlates with domestic patent application authorization and regional GDP, and we suppose that domestic patent application authorization is a quantifiable index indicating scientific and technological innovation, and that regional GDP is an excellent index indicating economic development. They play a great role in promoting the reduction of agricultural drought disasters at the beginning of the growth of scientific and technological innovation and economic development, while during the development of economy, science and technology reaches a certain level; for example, in our case, domestic patent application authorization equals 20,000, regional GDP equals 150 trillion yuan, and the promotion effect of economy, and science and technology on agricultural drought reduction will slow down. The responses of agricultural drought disasters to total sown area of crops and effective irrigation area are similar and can be basically divided into two stages. The first stage mainly corresponds to the stage when China's economy and society have not yet developed rapidly. Total sown area of crops and effective irrigation area are low, and the ability of science and technology to support drought disaster response is also poor. Therefore, with the increase of total sown area of crops and effective irrigation area, agricultural drought disasters are also increasing. The second stage mainly corresponds to the stage of rapid economic and social development in China. Total sown area of crops and effective irrigation area are rising, and the ability of

science and technology to support drought disaster response is also significantly enhanced. Therefore, with the increase of crop sown area, the growth rate of agricultural drought disasters slows down or even decreases. The implementation of irrigation measures also effectively mitigates drought events. With the increase of effective irrigation area, agricultural drought disasters are decreasing. Based on the aforementioned analysis, we can present the critical thresholds for agricultural drought disaster prevention and control in China: total sown area of crops >4 million hectares, effective irrigation area >1 million hectares, domestic patent application authorization >20,000, and regional GDP >150 trillion yuan.

Chen et al. (2018) explored the relationship between socioeconomic factors and grain vulnerability to drought in China covering the period of 1949–2015 using the method of locally weighted regression plots. Result shows that irrigated area rate and GDP in agriculture have a generally monotonous, linear, and positive relationship with grain sensitivity of drought, which means with the increase of irrigated area rate and GDP in agriculture, sensitive crop production to droughts decreases. This result is basically consistent with the effect of the second stage of effective irrigation area and regional GDP on agricultural drought disasters presented in our paper. Pogson et al. (2012) carried out sensitivity analysis based on a crop model in the UK, and results showed that with the increase of precipitation, relative crop yield is increasing with an S-curve shape. This result is close to the negative relationship between precipitation and agricultural drought disasters presented in our paper.

In addition, although other factors do not show significant influence on agricultural drought disasters, their ALE plots can still reflect their effect since ALE plots have already removed the interference of relevant factors. For example, as can be seen from **Figure 9** and **Supplementary Figure S5**, agricultural drought disasters is negatively correlated with F4 (water production modulus), but positively correlated with F8 (rural population).

Furthermore, it is worth noting that although R^2 of cross-validation and test for RF models are not high (<0.7), the relationship between influencing factors and target variables based on Monte Carlo simulation is relatively stable (**Figure 9** and **Supplementary Figure S5**). The purpose of this paper is to filter out the principal influencing factors of agricultural drought disasters, so the constructed RF model can fully meet this requirement. However, the accuracy of cross-validation and test is not high, which also shows that the input features cannot fully explain the variance of drought-affected area and drought-suffering area. It is necessary to further improve the input features to predict agricultural drought disasters in the future.

From the analysis of agricultural drought disasters in China, we can perceive that the north–south spatial pattern of agricultural drought is remarkable. Next, we will use unsupervised learning method to further refine agricultural drought disaster zoning. On this basis, based on RF algorithm, we will explore the differences of the key influencing factors of agricultural drought disasters in different zones. In addition, we will further analyze the occurrence and development process of agricultural drought disasters from the perspective of physical causes, and explore the relationship between meteorological drought, hydrological drought, and agricultural drought in near-natural areas and areas affected by human activities, so as to deepen the understanding of drought disasters and provide scientific basis for drought event monitoring, evaluation, early warning, and prediction.

6 CONCLUSION

In this study, we analyzed the spatial–temporal pattern and evolution characteristics of agricultural drought disasters in China based on the Mann–Kendall test, and applied Random Forest algorithm by integrating Gini importance, permutation feature importance, and accumulated local effects plot to quantify the role of natural and human factors on agricultural drought disasters. The constructed RF model can adequately meet the requirement to filter out the key factors of agricultural drought disasters, reveal the nonlinear response of agricultural drought disasters to the principal drivers, and identify the critical thresholds for agricultural drought disaster prevention and control.

The following conclusions can be drawn from this study:

1) Agricultural drought disaster has been fluctuating since 1950, which can be roughly divided into three stages: 1950–1980 with a significant upward trend, 1981–2000 with a non-significant upward trend, and 2001–2020 with a significant downward trend.
2) The spatial pattern of agricultural drought disaster tends to decrease in severity from north to south. Eight northern provinces including Inner Mongolia, Gansu, Jilin, Shaanxi, Hebei, Heilongjiang, Shandong, and Henan and three southern provinces including Yunnan, Hubei, and Hunan are greatly affected by drought. Among them, agricultural drought disasters in Shaanxi, Hebei, Shandong, Hunan, and Henan present a significant downward trend, while drought in Inner Mongolia, Gansu, Jilin, Heilongjiang, Yunnan, and Hubei shows a non-significant downward trend.
3) Total sown area of crops, precipitation, effective irrigation area, domestic patent application authorization, and regional GDP are the top 5 dominant factors influencing agricultural drought disasters.
4) Agricultural drought disasters have a negative correlation with precipitation, domestic patent application authorization, and regional GDP, and the nonlinear responses of agricultural drought disasters to total sown area of crops and effective irrigation area can be basically divided into two stages. In the first stage, with the increase of feature value, agricultural drought disaster is also increasing. In the second stage, with the increase of feature value, agricultural drought disaster is growing slow or just decreasing.
5) The critical thresholds for agricultural drought disaster prevention and control in China are as follows: total sown area of crops >4 million hectares, effective irrigation area >1

million hectares, domestic patent application authorization >20,000, and regional GDP >150 trillion yuan.

From these insights, we propose that more attention should be paid in northern China and three southern provinces including Yunnan, Hubei, and Hunan, and all provinces should further improve their response ability to heavy drought disasters. Furthermore, to improve the prediction ability of the RF model, it is necessary to integrate more knowledge about agricultural drought disasters. This implies that a better understanding of the role of factors on agricultural drought disasters and a better understanding of the development process of agricultural drought disasters complement each other.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

XD and GW contributed to conception and design of the study. XD organized the database. XD performed the statistical analysis. XD and GW wrote the first draft of the article. XD, JZ, and XL wrote sections of the article. XD, GW and HY contributed to article revision, read, and approved the submitted version.

FUNDING

This study was financially supported by the National Key R&D Program of China (2017YFA0605002, 2021YFC3201100), the National Natural Science Foundation of China (41830863, 52121006), the Belt and Road Fund on Water and Sustainability of the State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, China (2020nkzd01, 2021490211), and the Special Fund for Basic Scientific Research of Nanjing Hydraulic Research Institute (Y522002).

ACKNOWLEDGMENTS

Hersbach et al. (2019) was downloaded from the Copernicus Climate Change Service (C3S) Climate Data Store. The results contain modified Copernicus Climate Change Service information 2020. Neither the European Commission nor ECMWF is responsible for any use that may be made of the Copernicus information or data it contains. The authors would like to acknowledge editors and reviewers who proposed valuable comments that helped us to improve the article.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fenvs.2022.820615/full#supplementary-material

REFERENCES

Apley, D. W., and Zhu, J. Y. (2016). Visualizing the Effects of Predictor Variables in Black Box Supervised Learning Models. Available at: https://arxiv.org/abs/1612.08468.

Başakın, E. E., Ekmekcïoğlu, Ö., and Ozger, M. (2019). Drought Analysis with Machine Learning Methods. *Pamukkale J. Eng. Sci.* 25 (8), 985–991. doi:10.5505/pajes.2019.34392

Berghuijs, W. R., Hartmann, A., and Woods, R. A. (2016). Streamflow Sensitivity to Water Storage Changes across Europe, European Geosciences Union General Assembly. *Geophys. Res. Lett.* 43, 1980–1987. doi:10.1002/2016gl067927

Blauhut, V., Stahl, K., Stagge, J. H., Tallaksen, L. M., De Stefano, L., and Vogt, J. (2016). Estimating Drought Risk across Europe from Reported Drought Impacts, Drought Indices, and Vulnerability Factors. *Hydrol. Earth Syst. Sci.* 20 (7), 2779–2800. doi:10.5194/hess-20-2779-2016

Breiman, L. (2001). Random Forests. *Machine Learn.* 45 (1), 5–32. doi:10.1023/A:1010933404324

Cancelliere, A., Mauro, G. D., Bonaccorso, B., and Rossi, G. (2006). Drought Forecasting Using the Standardized Precipitation Index. *Water Resour. Manage.* 21 (5), 801–819. doi:10.1007/s11269-006-9062-y

Chen, H., Liang, Z., Liu, Y., Jiang, Q., and Xie, S. (2018). Effects of Drought and Flood on Crop Production in China across 1949-2015: Spatial Heterogeneity Analysis with Bayesian Hierarchical Modeling. *Nat. Hazards* 92, 525–541. doi:10.1007/s11069-018-3216-0

Cheng, J., Du, Z., Hu, J. L., and Li, C. S. (2017). Study on the Influencing Factors and Spatial Variation of Agricultural Drought Vulnerability: Based on Six Provinces and One City in the Middle and Lower Reaches of the Yangtze River. *Ecol. Economy* 33 (9), 188–194. in Chinese.

Dai, M., Huang, S., Huang, Q., Leng, G., Guo, Y., Wang, L., et al. (2020). Assessing Agricultural Drought Risk and its Dynamic Evolution Characteristics. *Agric. Water Manag.* 231, 106003. doi:10.1016/j.agwat.2020.106003

Dayal, K. S., Deo, R. C., and Apan, A. A. (2018). Spatio-temporal Drought Risk Mapping Approach and its Application in the Drought-Prone Region of South-East Queensland, Australia. *Nat. Hazards* 93 (2), 823–847. doi:10.1007/s11069-018-3326-8

Fienen, M. N., Nolan, B. T., Kauffman, L. J., and Feinstein, D. T. (2018). Metamodeling for Groundwater Age Forecasting in the Lake Michigan Basin. *Water Resour. Res.* 54 (7), 4750–4766. doi:10.1029/2017wr022387

Fisher, A., Rudin, C., and Dominici, F. (2019). All Models Are Wrong, but Many Are Useful: Learning a Variable's Importance by Studying an Entire Class of Prediction Models Simultaneously. *J. Mach. Learn. Res.* 20 (177), 1–81. arXiv: 1801.01489.

Grömping, U. (2020). "Model-Agnostic Effects Plots for Interpreting Machine Learning Models." Reports in Mathematics, Physics and Chemistry: Department II. Report 1/2020.

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., and Pedreschi, D. (2019). A Survey of Methods for Explaining Black Box Models. *ACM Comput. Surv.* 51 (5), 1–42. doi:10.1145/3236009

Gupta, H. V., and Nearing, G. S. (2015). Debates—the Future of Hydrological Sciences: A (Common) Path Forward? Using Models and Data to Learn: A Systems Theoretic Perspective on the Future of Hydrological Science. *Water Resour. Res.* 50, 5351–5359. doi:10.1002/2013wr015096

Han, Z., Huang, Q., Huang, S., Leng, G., Bai, Q., Liang, H., et al. (2021). Spatial-temporal Dynamics of Agricultural Drought in the Loess Plateau under a Changing Environment: Characteristics and Potential Influencing Factors. *Agric. Water Manag.* 244, 106540. doi:10.1016/j.agwat.2020.106540

Hao, Z., Yuan, X., Xia, Y., Hao, F., and Singh, V. P. (2017). An Overview of Drought Monitoring and Prediction Systems at Regional and Global Scales. *Bull. Am. Meteorol. Soc.* 98 (9), 1879–1896. doi:10.1175/bams-d-15-00149.1

Hengl, T., Nussbaum, M., Wright, M. N., Heuvelink, G. B. M., and Gräler, B. (2018). Random Forest as a Generic Framework for Predictive Modeling of Spatial and Spatio-Temporal Variables. *PeerJ.* 6, e5518. doi:10.7717/peerj.5518

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., et al. (2019). *ERA5 Monthly Averaged Data on Single Levels from 1979 to Present*. Copernicus Climate Change Service, Climate Data Store Reading. doi:10.24381/cds.f17050d7

Hong, X. (2017). Hydrological Drought Indices and Frequency Analysis Methods and Their Applications. Doctoral thesis. Wuhan (China): Wuhan University.

Huang, S., Huang, Q., Chang, J., Leng, G., and Xing, L. (2015). The Response of Agricultural Drought to Meteorological Drought and the Influencing Factors: A Case Study in the Wei River Basin, China. *Agric. Water Manag.* 159, 45–54. doi:10.1016/j.agwat.2015.05.023

Javadinejad, S., Dara, R., and Jafary, F. (2020). Analysis and Prioritization the Effective Factors on Increasing Farmers Resilience under Climate Change and Drought. *Agric. Res.* 10 (3), 497–513. doi:10.1007/s40003-020-00516-w

Ji, S. L., Li, J. F., Du, T. Y., and Li, B. (2019). Survey on Techniques, Applications and Security of Machine Learning Interpretability. *J. Comp. Res. Dev.* 56 (10), 2071–2096. in Chinese.

Karpatne, A., Atluri, G., Faghmous, J. H., Steinbach, M., Banerjee, A., Ganguly, A., et al. (2017). Theory-Guided Data Science: A New Paradigm for Scientific Discovery from Data. *IEEE Trans. Knowledge Data Eng.* 29, 2318–2331. doi:10.1109/tkde.2017.2720168

Koch, J., Stisen, S., Refsgaard, J. C., Ernstsen, V., Jakobsen, P. R., and Højberg, A. L. (2019). Modeling Depth of the Redox Interface at High Resolution at National Scale Using Random Forest and Residual Gaussian Simulation. *Water Resour. Res.* 55 (2), 1451–1469. doi:10.1029/2018wr023939

Kohavi, R., and Provost, F. (1998). Special Issue on Applications of Machine Learning and the Knowledge Discovery Process. *Machine Learn.* 30, 127–132.

Konapala, G., and Mishra, A. K. (2016). Three-parameter-based Streamflow Elasticity Model: Application to MOPEX Basins in the USA at Annual and Seasonal Scales. *Hydrol. Earth Syst. Sci.* 20 (6), 2545–2556. doi:10.5194/hess-20-2545-2016

Li, R., Xin, J. F., and Yang, Y. M. (2019). Analysis on the Temporal and Spatial Change of Drought in Northeast China from 1949 to 2017. *Water Resour. Hydropower Eng.* 50 (S2), 1–6. doi:10.13928/j.cnki.wrahe.2019.S2.001

Liao, Y. M., and Zhang, C. J. (2017). Spatio-temporal Distribution Characteristics and Disaster Change of Drought in China Based on Meteorological Drought Composite index. *Meteorol. Monthly* 43 (11), 1402–1409. in Chinese.

Liu, W. D., Tang, Z. P., Xia, Y., Han, M. Y., and Jiang, W. B. (2019). Identifying the Key Factors Influencing Chinese Carbon Intensity Using Machine Learning, the Random forest Algorithm, and Evolutionary Analysis. *Acta Geographica Sinica* 74 (12), 2592–2603. in Chinese.

Liu, X., Zhu, Z., Liu, X., and Yu, M. (2021). Thresholds of Key Disaster-Inducing Factors and Drought Simulation in the Xilinguole Grassland. *Ecol. Inform.* 64, 101380. doi:10.1016/j.ecoinf.2021.101380

Liu, Z., Li, Q., Nguyen, L., and Xu, G. (2018). Comparing Machine-Learning Models for Drought Forecasting in Vietnam's Cai River Basin. *Pol. J. Environ. Stud.* 27 (6), 2633–2646. doi:10.15244/pjoes/80866

Lv, J., Gao, H., and Sun, H. (2011). Analysis on the Characteristics and Causes of Drought Disasters in China since the 21st century. *China Flood Drought Manag.* 21 (05), 38–43. in Chinese. doi:10.16867/j.cnki.cfdm.2011.05.014

Ministry of water resources of China (2019). *China Flood and Drought Disaster Bulletin, 2018*. Beijing: China Water Power Press. in Chinese.

Molnar, C. (2021). Interpretable Machine Learning. Available at: https://christophm.github.io/interpretable-ml-book/.

Ni, S. H., Gu, Y., Peng, Y. J., Liu, J. N., and Wang, H. L. (2019). Spatio-temporal Pattern and Evolution Trend of Drought Disaster in China in Recent Seventy Years. *J. Nat. Disasters* 28 (06), 176–181. in Chinese. doi:10.13577/j.jnd.2019.0619

Nourani, V., Baghanam, A. H., Adamowski, J. K., and Kisi, O. (2014). Applications of Hybrid Wavelet-Artificial Intelligence Models in Hydrology: A Review. *J. Hydrol.* 514, 358–377. doi:10.1016/j.jhydrol.2014.03.057

Orth, R., and Destouni, G. (2018). Drought Reduces Blue-Water Fluxes More Strongly Than green-water Fluxes in Europe. *Nat. Commun.* 9 (1), 3602. doi:10.1038/s41467-018-06013-7

Pang, S. F., Wei, W., Guo, Z. C., Zhang, J., Jie, B. B., et al. (2019). Agricultural Drought Characteristics and its Influencing Factors in Gansu Province Based on TVDI. *Chin. J. Ecol.* 38 (06), 1849–1860. in Chinese. doi:10.13292/j.1000-4890.201906.035

Pogson, M., Hastings, A., and Smith, P. (2012). Sensitivity of Crop Model Predictions to Entire Meteorological and Soil Input Datasets Highlights Vulnerability to Drought. *Environ. Model. Softw.* 29 (1), 37–43. doi:10.1016/j.envsoft.2011.10.008

Probst, P., and Boulesteix, A. L. (2017). To Tune or Not to Tunethe Number of Trees in Random Forest?. *arXiv preprint*. arXiv:1705.05654.

Qian, X., Liang, L., Shen, Q., Sun, Q., Zhang, L., Liu, Z., et al. (2016). Drought Trends Based on the VCI and its Correlation with Climate Factors in the Agricultural Areas of China from 1982 to 2010. *Environ. Monit. Assess.* 188 (11), 639. doi:10.1007/s10661-016-5657-9

Raghavendra. N, S. S., and Deka, P. C. (2014). Support Vector Machine Applications in the Field of Hydrology: A Review. *Appl. Soft Comput.* 19, 372–386. doi:10.1016/j.asoc.2014.02.002

Schwalm, C. R., Anderegg, W. R. L., Michalak, A. M., Fisher, J. B., Biondi, F., Koch, G., et al. (2017). Global Patterns of Drought Recovery. *Nature* 548 (7666), 202–205. doi:10.1038/nature23021

Shah, D., and Mishra, V. (2020). Integrated Drought Index (IDI) for Drought Monitoring and Assessment in India. *Water Resour. Res.* 56, e2019WR026284. doi:10.1029/2019wr026284

Shamshirband, S., Hashemi, S., Salimi, H., Samadianfard, S., Asadi, E., Shadkani, S., et al. (2020). Predicting Standardized Streamflow index for Hydrological Drought Using Machine Learning Models. *Eng. Appl. Comput. Fluid Mech.* 14 (1), 339–350. doi:10.1080/19942060.2020.1715844

Su, B. D., Sun, H. M., Li, X. C., Li, Z. J., Zhang, J. P., Wang, Y. J., et al. (2020). Impact of Climate Change on Terrestrial Water Cycle in China. *China J. Atmosphere Sci.* 43 (06), 1096–1105. in Chinese. doi:10.13878/j.cnki.dqkxxb.20201014001

Tian, Q., Lu, J., and Chen, X. (2022). A Novel Comprehensive Agricultural Drought index Reflecting Time Lag of Soil Moisture to Meteorology: A Case Study in the Yangtze River basin, China. *Catena* 209, 105804. doi:10.1016/j.catena.2021.105804

Wang, W. D., Sun, L., Dong, M., Li, F. X., and Liu, M. (2021a). Analysis of Drought Characteristics of Crop Growing Season Based on SPEI in Northeast China. *J. Anhui Agric. Sci.* 49 (03), 64–68. in Chinese.

Wang, W. D., Sun, L., Pei, Z. Y., Ma, S. J., Chen, Y. Y., Sun, J. Y., et al. (2021b). Effect of Growing Season Drought and Flood on Yield of spring maize in Three Northeast Provinces of China. *Smart Agric.* 3 (02), 126–137. in Chinese.

West, H., Quinn, N., and Horswell, M. (2019). Remote Sensing for Drought Monitoring & Impact Assessment: Progress, Past Challenges and Future Opportunities. *Remote Sensing Environ.* 232, 111291. doi:10.1016/j.rse.2019.111291

Wu, B., Ma, Z., and Yan, N. (2020). Agricultural Drought Mitigating Indices Derived from the Changes in Drought Characteristics. *Remote Sensing Environ.* 244, 111813. doi:10.1016/j.rse.2020.111813

Wu, C. G., Bai, L., Bai, X., Jin, J. L., and Jiang, S. M. (2018). Risk Assessment and Division Model for Regional Drought Disaster Based on Cloud Model and Bootstrap Method. *J. Nat. Resour.* 33 (04), 684–695. in Chinese.

Wu, J., Lin, X., Wang, M., Peng, J., and Tu, Y. (2017). Assessing Agricultural Drought Vulnerability by a VSD Model: A Case Study in Yunnan Province, China. *Sustainability* 9 (6), 918. doi:10.3390/su9060918

Yan, H., Wang, S.-Q., Wang, J.-B., Lu, H.-Q., Guo, A.-H., Zhu, Z.-C., et al. (2016). Assessing Spatiotemporal Variation of Drought in China and its Impact on Agriculture during 1982-2011 by Using PDSI Indices and Agriculture Drought Survey Data. *J. Geophys. Res. Atmos.* 121 (5), 2283–2298. doi:10.1002/2015jd024285

Yang, Q., Luo, G., and Gao, C. (2020). Research Progress of Agricultural Drought from the Perspective of Geography. *J. North China Univ. Water Resour. Electric Power (Natural Sci. Edition)* 41 (01), 27–34. in Chinese. doi:10.19760/j.ncwu.zk.2020004

Zhang, D., Cao, W., and Qi, B. (2021). Identifying Influencing Factors of Regional Agricultural Drought Vulnerability Based on PSR-TGRC Method. *Math. Probl. Eng.* 2021, 1–13. doi:10.1155/2021/9933152

Zhang, S. F., Su, Y. S., Song, D. D., Zhang, Y. Y., Song, H. Z., Gu, Y., et al. (2008). *The Drought in Chinese History 1949-2000*. Nanjing: Hohai University Press. in Chinese.

Zhao, J., Zhang, Q., Zhu, X., Shen, Z., and Yu, H. (2020). Drought Risk Assessment in China: Evaluation Framework and Influencing Factors. *Geogr. Sustainability* 1 (3), 220–228. doi:10.1016/j.geosus.2020.06.005

Zhou, Z., Shi, H., Fu, Q., Ding, Y., Li, T., and Liu, S. (2021). Investigating the Propagation from Meteorological to Hydrological Drought by Introducing the Nonlinear Dependence with Directed Information Transfer index. *Water Res.* 57, e2021WR030028. doi:10.1029/2021WR030028

Zobeidi, T., Yazdanpanah, M., Komendantova, N., Sieber, S., and Löhr, K. (2021). Factors Affecting Smallholder Farmers' Technical and Non-technical Adaptation Responses to Drought in Iran. *J. Environ. Manag.* 298, 113552. doi:10.1016/j.jenvman.2021.113552

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Environmental Science

ORIGINAL RESEARCH
published: 01 June 2022
doi: 10.3389/fenvs.2022.812351

The Possible Incoming Runoff Under Extreme Rainfall Event in the Fenhe River Basin

Shengqi Jian[1], Changyan Yin[1], Yafei Wang[1], Xin Yu[2] and Yong Li[2]*

[1]College of Water Conservancy Science and Engineering, Zhengzhou University, Zhengzhou, China, [2]Yellow River Institute of Hydraulic Research, Yellow River Conservancy Commission, Zhengzhou, China

Prediction of runoff is of great significance for the sustainable utilization of water resources and flood control and disaster reduction in the basin. In this study, a method for predicting the runoff caused by extreme sub-rainfall events was constructed based on the identification of extreme rainfall events, Mann–Kendall Test, R/S analysis, and regression analysis. The method was applied to the Jingle sub-basin, and the results showed that the extreme precipitation in this basin will maintain a slight rising trend in the future, assuming that the climate and underlying conditions remain the same as they were in the current scenario. There is a more stable correlational relationship between rainfall characteristic factors and runoff in extreme rainfall events. The extreme precipitation of 1–5 consecutive days under the 100a return period designed by the hydrological frequency method is 38.74, 60.01, 66.00, 71.44, and 73.69 mm, respectively, and the possible runoff predicted by the four empirical formulas is 1295–2495, 2108–4935, 2408–5801, and 3051–7062 × 10^4 m^3, respectively. The rainfall designed by the hydrological frequency combination method is 203.64 mm, and the possible runoff predicted by the four empirical formulas is 2.8–5.3 × 10^8 m^3. This study can provide a new reference for predicting the possible incoming runoff under extreme sub-rainfall events.

Keywords: extreme rainfall event, rainfall event, jingle sub-basin, runoff prediction, regression analysis

Edited by:
Chuanfu Zang,
South China Normal University, China

Reviewed by:
Peng Shouzhang,
Northwest A&F University, China
Huiliang Wang,
Zhengzhou University, China

****Correspondence:***
Shengqi Jian
jiansq@zzu.edu.cn

Specialty section:
This article was submitted to Environmental Informatics and Remote Sensing, a section of the journal Frontiers in Environmental Science

Received: *15 November 2021*
Accepted: *20 April 2022*
Published: *01 June 2022*

Citation:
Jian S, Yin C, Wang Y, Yu X and Li Y (2022) The Possible Incoming Runoff Under Extreme Rainfall Event in the Fenhe River Basin. Front. Environ. Sci. 10:812351. doi: 10.3389/fenvs.2022.812351

1 INTRODUCTION

Analysis and prediction of runoff characteristics can provide a reasonable basis for rational regulation and optimal allocation of water resources, water resource protection and planning, and effective management of water resources. Under the combined influence of climate change and human activities, the relationship between rainfall and runoff presented uncertainty, multiple time scales, randomness, chaos, weak dependent, highly complex nonlinear, and non-stationary characteristics (Galelli and Castelletti, 2013; Zhang et al., 2018). These change characteristics proposed great challenges for the prediction of runoff. The commonly used methods of runoff prediction include the process-driven model and data-driven model. The process-driven model focuses on the process of flow generation and concentration to simulate the runoff process from the perspective of hydrological principles, such as HEC-HMS (Teng et al., 2017), SWAT (Wu et al., 2019), Xin'anjiang (Hao et al., 2018), and MIKE-SHE (Qi et al., 2021). From the perspective of data analysis and mining, the data-driven model analyzes the flow process and its influencing factors, combines the mathematical–statistical relationship between data input and output, and constructs a model for runoff prediction, such as the regression model (Qamar et al., 2016; Visessri and McIntyre, 2016), artificial neural networks (Seckin et al., 2013; Gökbulak et al., 2015), and support vector

machine (SVM) (Li et al., 2013; Wang et al., 2014). Most of the researchers used effective methods and obtained ideal results for runoff prediction and rainfall–runoff process simulation. But, only a few studies focus on runoff prediction on a small time-scale, especially on the scale of the sub-rainfall event.

Global warming caused by climate change has become one of the most serious environmental challenges facing the world (Sun et al., 2015). Climate change will pose a severe threat to the global and regional ecological environment, among which the impact of extreme rainfall on nature and society is far greater than that of others (Manfreda et al., 2018; Barlow et al., 2019). An increasing amount of research has found that extreme precipitation characteristics have changed around the world (Westra et al., 2013; Tan et al., 2021). The relationship between rainfall characteristics and runoff generation varied within different rainfall levels and intensities (de Lima et al., 2009; Ran et al., 2012), so numerous researchers have investigated the characteristics of water and sediment under extreme rainfall separately (Liu et al., 2019; Dang et al., 2020; Zhao et al., 2020). Liu et al. (2019) took Hekou to the Tongguan section in the middle reaches of the Yellow River as the research area and analyzed the possible annual sediment and runoff yield under extreme rainstorm conditions by the SWAT model. Dang et al. (2020), based on historical rainfall sediment records, studied the relationship between precipitation and sediment yield and predicted the annual sediment under the design extreme precipitation scenario. Therefore, from the perspective of the sub-rainfall event, this study predicts the possible runoff under extreme sub-rainfall events.

As mentioned above, the main purpose of this study is to predict the runoff under extreme sub-rainfall events. We used the Jingle sub-basin as the study area, and the following steps were carried out. 1) Based on the definition of the extreme precipitation index, the selection criteria of sub-extreme rainfall event in this study were defined. 2) The trend and consistency of extreme rainfall were analyzed by the MK test and R/S analysis, respectively. 3) Different rainfall factors were selected to establish several rainfall–runoff empirical formulas by regression analysis. 4) Then, the empirical formulas combined with designed extreme rainfall scenarios were used to predict the possible incoming runoff under extreme rainfall events.

2 MATERIALS AND METHODS

2.1 Study Area

Fenhe River is located in the middle reaches of the Yellow River and is the second-largest tributary of the Yellow River, with a total length of 716 km and a drainage area of 39,471 km^2 (110°30'–113°32'E; 35°20'–39°00'N). The climate in this basin differs significantly as a result of the complex atmospheric and monsoon circulation in the mid-latitude zone. The annual average evaporation is 1120 mm, and the average annual precipitation is 503 mm. The average annual runoff is 2.28×10^9 m^3. The spatial distribution of rainfall is uneven, showing a decreasing trend from south to north. Approximately 60–80% of the annual precipitation falls in the form of heavy rain and is temporally concentrated between June and September. Due to the gentle river course and the influx of sediment carried by many tributaries, flood disasters occur frequently in the Basin. In the past 100 years, there were more than 20 record floods in the Fenhe River, roughly once every 5 years. After 1949, there were five relatively large-scale floods in the middle and lower reaches of the Fenhe River. Among them, the magnitude of the flood that occurred on 21 August 1982 was the largest, with a peak discharge of 1420 m^3/s, which caused huge economic losses and casualties. It can be seen that flood disasters have always been a vital and prominent issue in the basin.

The Jingle sub-basin (**Figure 1**) lies in the upper reaches of Fenhe River, controlling approximately 1/3 of the area of the upper reaches of Fenhe River. The area of the basin is 2799 km^2, and the length of the river is about 83.9 km. The Jingle Hydrological Station is a control station for the Jingle sub-basin. The average annual precipitation is 497.85 mm, and the monthly average temperature is 4–13°C. Construction lands substantially expanded after 2000 with the rapid development of the economy and urbanization across the whole basin. Concomitantly, vegetation coverage has increased in the basin due to the implementation of the national level "Soil and Water Conservation" program for 20 years, beginning in 1988, but the mainland use types in this area are still woodland and arable land.

2.2 Data Collection and Processing

The measured daily precipitation data of five hydrological stations in the Fenhe River Basin (**Figure 1B**) from 1960 to 2019 are obtained from the National Meteorological Science Data Center (http://data.cma.cn/), and some missing data were interpolated reasonably by the hydrological analogy method and linear interpolation method. The measured rainfall–runoff process data (1971–2018) of the hydrological station and meteorological station of the Jingle sub-basin were systematically unified into a 1-h time step (**Figure 1C**). A total of 103 rainfall events were selected according to the principle of no precipitation within 24 h. The area-averaged rainfall of sub-rainfall events was obtained *via* the Thiessen polygon method, and the rainfall characteristic variables (volume, period intensity, and duration) were counted (**Figure 2**).

2.3 Methods

2.3.1 Definition of Extreme Rainfall Event

The IPCC provides little information on the definition of the sub-extreme rainfall event, considering the long duration of rainfall events in the study area, and instead of fixed absolute threshold values, relative threshold values of rainfall corresponding to the 95^{th} percentiles were proposed to represent the extreme rainfall event. For rainfall events with a duration less than 24 h, the criterion to judge whether it is an extreme rainfall event is its total rainfall beyond the 95^{th} percentile threshold of daily precipitation; for rainfall events lasting 24–48 h, the criterion to judge whether it is an extreme rainfall event is its total rainfall beyond the 95^{th} percentile threshold of rainfall during two consecutive days; similarly, for rainfall events lasting 48–72 h, 72–96 h, and more than 96 h, the judgment criteria are that the total rainfall of them beyond the 95^{th} percentile threshold of

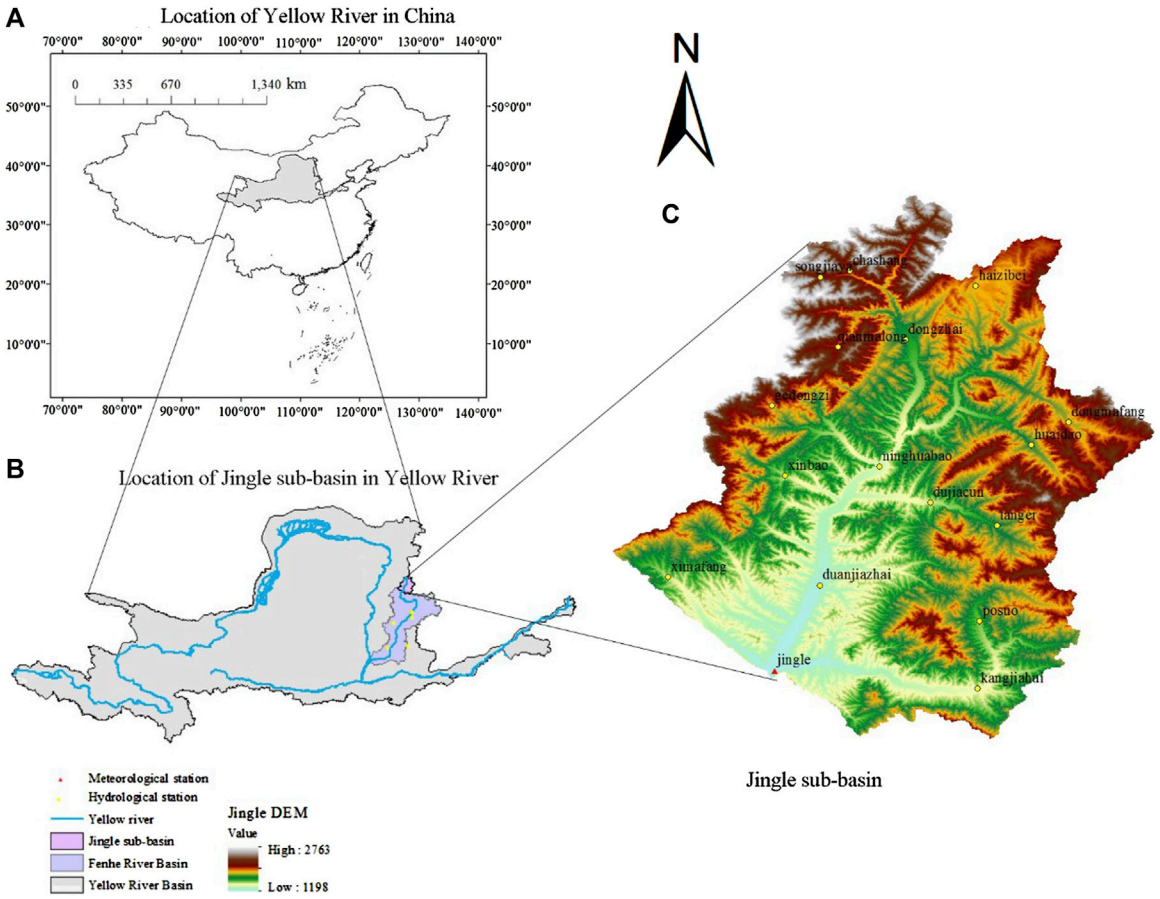


FIGURE 1 | Topographic map of the Jingle sub-basin. **(A)** Description of Yellow River in China; **(B)** Description of the Jingle sub-basin in the Yellow River Basin; **(C)** Control catchment of the Jingle station in Fenhe River and distribution of rainfall gage stations.

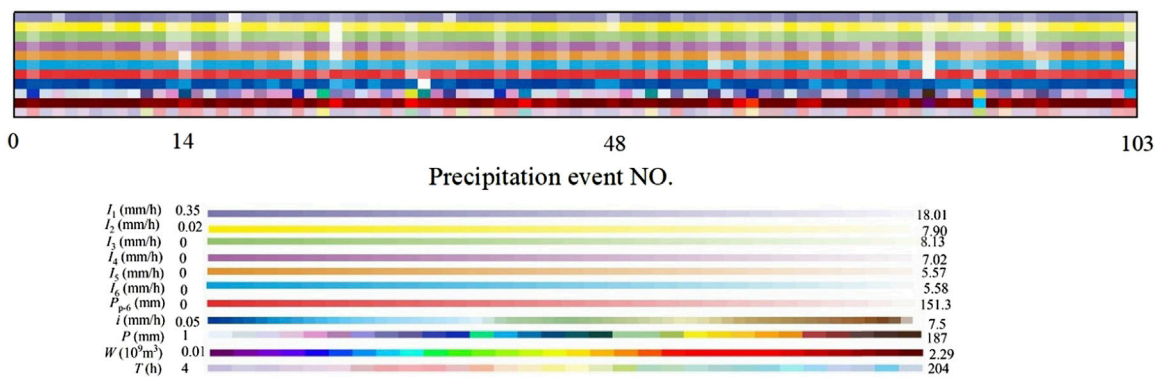


FIGURE 2 | Values of sub-rainfall event characteristic factors, 1–103 are the numbers of sub-rainfall events. I_1 is the maximum 1 h rain intensity; I_{2-1} is the second-largest hourly rainfall intensity; Similarly, I_{3-2}, I_{4-3}, I_{5-4}, and I_{6-5} are the 3rd, 4th, 5th, and 6th largest hourly rainfall intensity, respectively; $P_{p\text{-}6}$ is the rainfall minus the maximum rainfall of 6 h; P is the rainfall; W is the flood amount yielded by sub-rainfall; T is the sub-rainfall duration.

rainfall during three consecutive days, rainfall during four consecutive days, and rainfall during five consecutive days. The specific implementation steps are as follows: according to the collected daily precipitation data of five stations in the Fenhe River Basin from 1960 to 2019. The kriging interpolation method was adopted to obtain the daily precipitation data of the Jingle sub-basin from 1960 to 2019. When the daily precipitation is greater than 0.1 mm, it is considered to be a rainy day. The series

TABLE 1 | Classification of the Hurst index.

Grades	Range of *H* index	Anti-persistence	Grades	Range of *H* index	Persistence
–1	$0.45 \le H < 0.5$	Very weak	1	$0.50 < H \le 0.55$	Very weak
–2	$0.35 \le H < 0.45$	Relatively weak	2	$0.55 < H \le 0.65$	Relatively weak
–3	$0.25 \le H < 0.35$	Relatively strong	3	$0.65 < H \le 0.75$	Relatively strong
–4	$0.20 \le H < 0.25$	Strong	4	$0.75 < H \le 0.80$	Strong
–5	$0.00 \le H < 0.20$	Very strong	5	$0.80 < H \le 1.00$	Very strong

of rainy days in each year is arranged in ascending order of size to get $x_1, x_2, \ldots, x_n$ and then the 95th percentile threshold of daily precipitation (1 day) each year is as follows:

$$x = (1 - a)x_j + ax_{j+1}, \quad (1)$$

where j is the sequence number of daily rainfall arranged in ascending order, $j = Int[P(n + 1)]$, $Int[]$ is the integral function; P is the corresponding percentile; and a is the weight coefficient, $a = P(n + 1) - j$

Similarly, the 95th percentile threshold of rainfall during two consecutive days, rainfall during three consecutive days, rainfall during four consecutive days, and rainfall during five consecutive days of each year can be obtained. In this study, the series comprising 95th percentile threshold of daily precipitation of each year from 1960 to 2019 was called C1 day, the next was called C2 day, and so on until C5 day. The 60 year' average values of C1 day, C2 day, C3 day, C4 day, and C5 day are defined as the thresholds of extreme rainfall events with different durations.

2.3.2 Trend Analysis of Extreme Rainfall

Since the nonparametric MK test does not directly give the amplitude of the upward or downward trend, a simple linear regression method combined with the MK test was used to detect the trend significance and mutation test of the series (C1 day, C2 day, C3 day, C4 day, and C5 day), and their statistical significances were detected at the 0.05 level (Shi et al., 2016b; Shi et al., 2018).

R/S analysis was used to calculate the Hurst values of the series (C1 day, C2 day, C3 day, C4 day, and C5 day) in the Jingle sub-basin during 1960–2019, and the future trend of change can be evaluated according to the value of H ($0 < H < 1$) as follows: 1) when $H = 0.5$, the series is independent of each other, that is, future changes are independent of historical changes; 2) when $0 < H < 0.5$, the series of each element has anti-persistence, that is, the future changes will be opposite to past changes; 3) when $0.5 < H < 1$, future changes are consistent with past changes (Wu et al., 2021). Anti-persistence or persistence can be divided into five grades according to the strength from weak to strong (**Table 1**).

2.3.3 Empirical Formula Building

As the direct source of runoff, rainfall has a close relationship with runoff, which is manifested in the linear or nonlinear relationship between rainfall characteristic factors and runoff. In this study, the rainfall–runoff empirical formula between multiple rainfall characteristic factors and runoff was established by curve regression and multiple stepwise linear regression analysis. The specific steps of curve regression analysis are as follows: 1) the sub-rainfall prediction factor is taken as the abscissa and the sub-runoff as the ordinate and the scatter diagram of two variables is drawn; 2) the regression models between predictive variables and dependent variables were established through linear regression, binomial regression, power function regression, exponential function regression, and logarithmic function. 3) Finally, the decisive coefficient is taken as the criterion to choose the best rainfall–runoff empirical model.

2.3.3.1 The Method of Sub-Rainfall Fitting

Rainfall is the sum of the amount of rainfall in a certain period of time, which is the direct source of the amount of flood. The sub-rainfall fitting method is to use the total rainfall of an extreme rainfall event as the independent variable and the flood amount as the dependent variable to establish rainfall–runoff empirical **formula 1** through curve regression analysis.

2.3.3.2 The Method of Rainfall Factors Combination Fitting

The rainfall factors leading to soil erosion include rainfall and rainfall intensity. In order to consider the impact of the two factors on the sub-rainfall–runoff event at the same time, the product (PI_1) of rainfall (P) and the maximum 1 h rainfall intensity (I_1) of the sub-rainfall event was used as the prediction factor to establish the rainfall–runoff empirical **formula 2** by curve regression.

2.3.3.3 The Method of Sub-Rainfall Time-Segment Rainfall Fitting

In the process of precipitation, as the amount of rainfall varies, the influence of rainfall on water yield is different in various periods. Therefore, considering the influence of rainfall in different periods on the water yield, the empirical **formula 3** which describes the relationship between seven sub-periods (P_1, P_{2-1}, $P_{3-2} \ldots P_{p-6)}$ and runoff of sub-rainfall events was established by multiple stepwise linear regression analysis.

2.3.3.4 The Method of Upper Envelop

To consider complex and adverse situations in depth, the point data on the upper edge of the rainfall–runoff figure were selected to plot the rainfall–runoff upper envelop and fit the upper envelop relation through curve regression to get empirical **formula 4**.

2.3.4 Extreme Rainfall Event Scenario Setting and Possible Incoming Flood Prediction

Based on historical extreme precipitation records in the basin, two extreme rainfall scenarios were constructed in this study, and

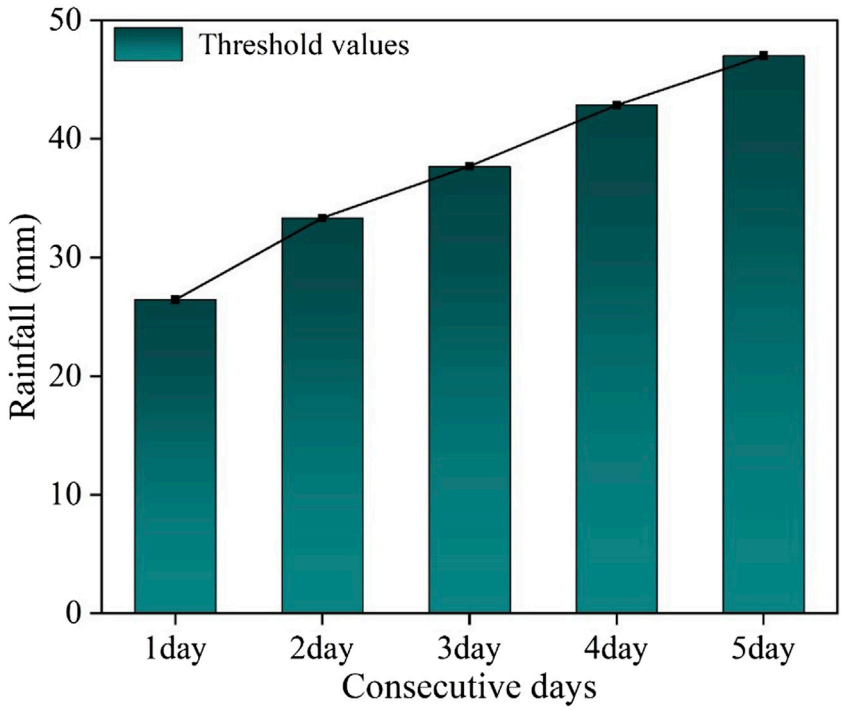


FIGURE 3 | 95th percentile threshold of extreme rainfall events with different durations.

the possible runoff under extreme precipitation scenarios was predicted by the established empirical model. The details of the scenario setting are as follows:

2.3.4.1 The Method of Hydrologic Frequency Analysis

The occurrence of rainfall events in the Jingle sub-basin is random, so it can be analyzed and calculated by the frequency analysis method.

The normal, gamma, gen gamma, log gamma, generalized extreme value, Gumbel max, and Weibull distribution were used to fit the C1 day, C2 day, C3 day, C4 day, and C5 day in the Jingle sub-basin. These functions were fitted using the maximum likelihood estimation method. In our study, Kolmogorov–Smirnov (K-S) analysis was selected to test if the data follow one of the specified distributions well. The hypothesis is evaluated at the 0.05 significance level. The distribution that passes the significance level test and has the smallest statistical value was selected as the optimal distribution. From the historical rainfall event record, the rainfall events with similar duration were taken as typical rainfall events to obtain the design rainfall process under different duration and different return periods by the fragment method.

2.3.4.2 The Method of Historical Measured Extreme Rainfall Combination

It is assumed that the rainfall extreme values of seven sub-periods (P_1, P_{2-1}, P_{3-2}, P_{4-3}, P_{5-4}, P_{6-5}, and P_{p-6}) in the historical measured rainfall events are all combined into one rainfall event so that the rainfall intensity and concentration degree of the composite rainfall were larger.

3 RESULTS AND DISCUSSION

3.1 Definition of Extreme Rainfall Event

According to the definition, the mean value (1960–2019) of the C1 day, C2 day, C3 day, C4 day, and C5 day series was obtained (**Figure 3**). That is, for rainfall events with a duration of less than 24 h, the value to judge whether it is an extreme rainfall event is the total rainfall of it beyond 24.66 mm; for rainfall events lasting 24–48 h, 48–72 h, 72–96 h, and more than 96 h, the judgment criteria are the total rainfall of them beyond 33.33, 37.68, 42.84, and 47.00 mm, respectively.

3.2 Trend Analysis of Extreme Rainfall

Simple linear regression and MK trend analysis show that in the past 55 years, the series of C1 day, C2 day, C3 day, C4 day, and C5 day increased slightly with the rate of 0.5 mm/10a, 0.7 mm/10a, 0.4 mm/10a, 1.1 mm/10a, and 0.6 mm/10a, respectively (**Figure 4**) (**Table 2**). The series of C1 day, C2 day, C3 day, C4 day, and C5 day generally had mutation points between 1964 and 2008, but the mutation was not significant (**Figure 5**).

To predict future increasing or decreasing trends, we used R/S analysis to calculate the *H* values of the series of C1 day, C2 day, C3 day, C4 day, and C5 day in the Jingle sub-basin during 1960–2019. The R/S analysis results of the series are shown in **Figure 6**; it can be seen that the Hurst indices of the series are 0.67, 0.61, 0.71, 0.57, and 0.54, indicating that the persistence of the five series is in the one-to-third intensity grade. The trends of them have a relatively weak persistence.

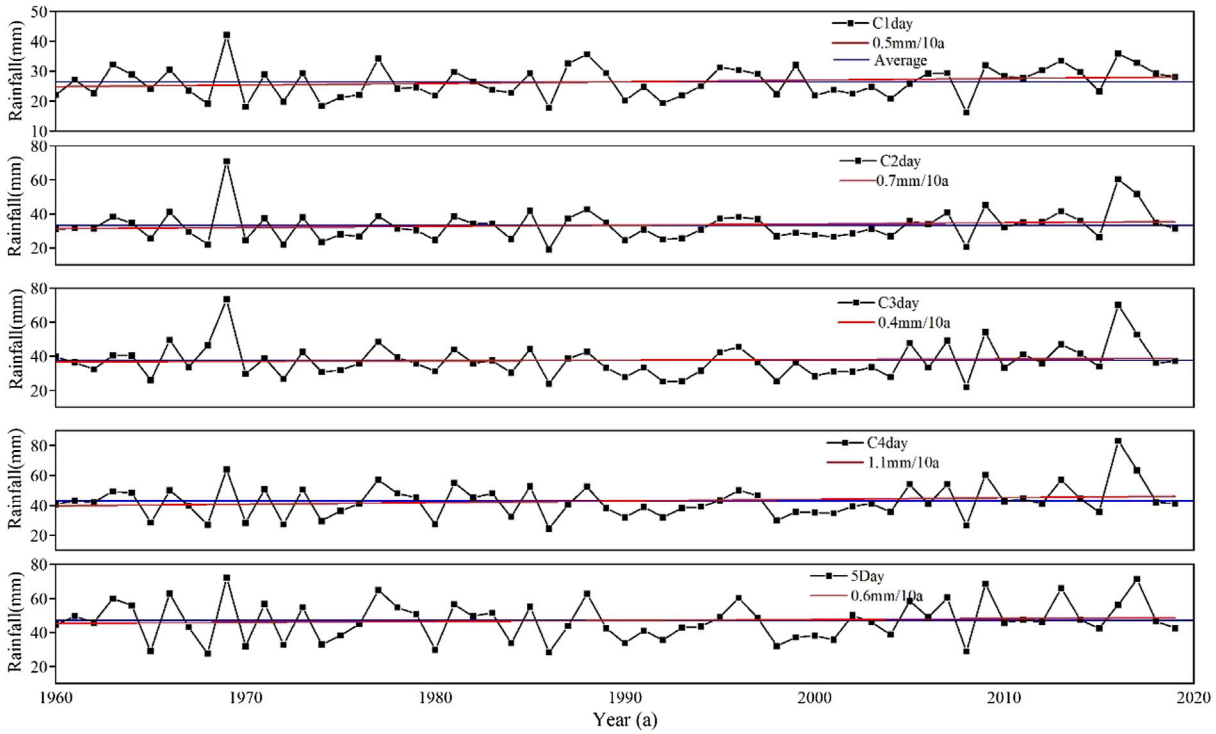


FIGURE 4 | Results of simple linear regression analysis.

TABLE 2 | Result of MK trend analysis.

Series	$Z_{1-\alpha}$	Z	Trends
C1 day	1.96	1.81	Increased slightly
C2 day	1.96	1.31	Increased slightly
C3 day	1.96	0.37	Increased slightly
C4 day	1.96	0.91	Increased slightly
C5 day	1.96	0.62	Increased slightly

Combining the results of the MK trend analysis and R/S analysis, the trend shows slight increments in extreme rainfall of the Jingle sub-basin, with a high likelihood of the trend continuing in the future, assuming the climate and underlying surface conditions remain the same as they were in the current scenario. This indicates that the calculated formula can be fitted using the rainfall–runoff series from 1971 to 2018 to predict the possible incoming flood amount in the future under extreme rainfall events.

3.3 Rainfall–Runoff Empirical Model

3.3.1 The Method of Sub-Rainfall Fitting

For 103 rainfall events and screened extreme rainfall events in the Jingle sub-basin from 1971 to 2018, Pearson's correlation coefficients between the total rainfall of an event and runoff are 0.77 and 0.83, respectively, indicating that rainfall is more related to runoff in extreme rainfall events. The binomial regression analysis in curve regression can better describe the relationship between rainstorm and runoff. The binomial coefficient in the empirical formula of all rainfall events is smaller than that of extreme rainfall events, indicating that per unit of rainfall yields more runoff in extreme rainfall events (**Figure 7**).

3.3.2 The Method of Rainfall Factor Combination Fitting

The rainfall factor combination (PI_1) can express the combined effect of rainfall and intensity on water and sediment yield in the watershed. Regression analysis shows that the binomial regression ideally simulates the rainstorm–runoff relationship, the binomial coefficients in the empirical formula of all rainfall events are smaller than those of extreme rainfall events, and it shows that runoff is more sensitive to the change of PI_1 in extreme rainfall events (**Figure 8**). In the study area, under the extreme rainfall event, rainfall factor combination fitting methods are better suited to two types of rainfall: type Ⅰ storms were characterized by a long duration and heavy rainfall amount, whereas type Ⅱ storms had a higher concentration and higher intensities. The rainfall types of the deviation point data on 28 July 1982, 8 July 1985, and 19 August 1973 are intermediate between the first two types, and the rainfall and intensity are lower than both, indicating that the fitting method fits the extreme scenarios better.

3.3.3 The Method of Sub-Rainfall Time-Segment Rainfall Fitting

When using the rainfall in seven sub-periods (P_1, P_{2-1}, P_{3-2}, P_{4-3}, P_{5-4}, P_{6-5}, and P_{p-6}) as input variables for stepwise regression analysis, these variables are independent of each other. The

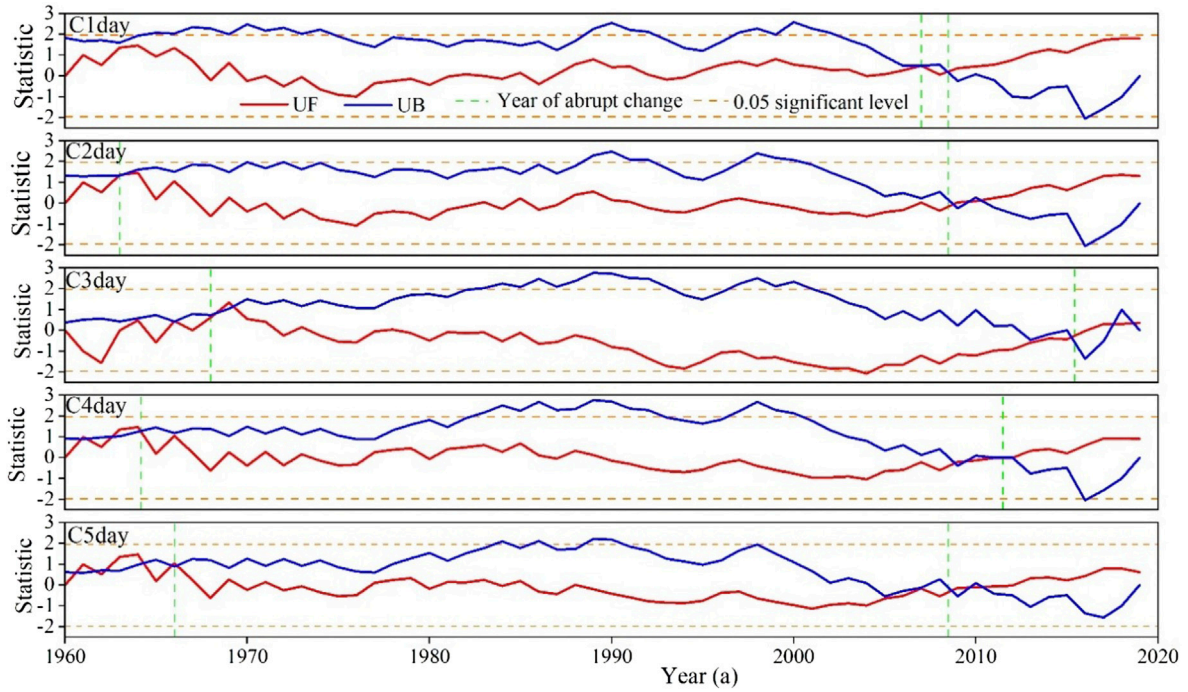


FIGURE 5 | Mann–Kendall (MK) mutation test for the series.

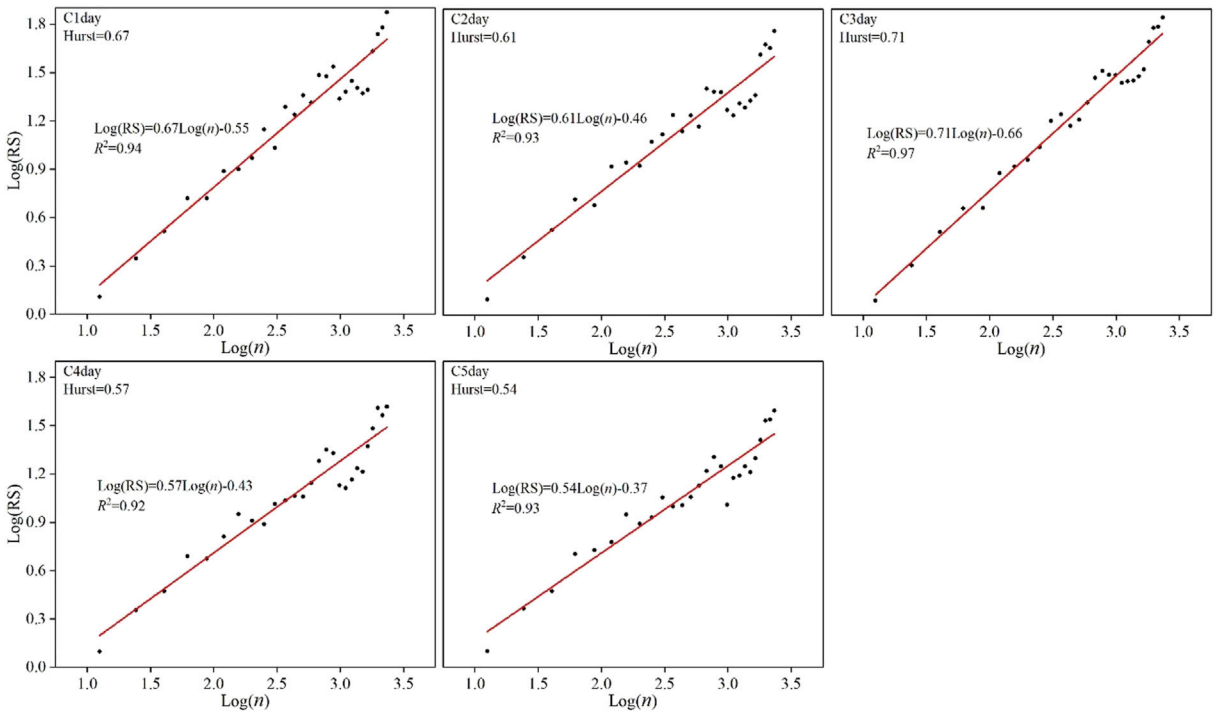


FIGURE 6 | Hurst long-range forecasting for the time series.

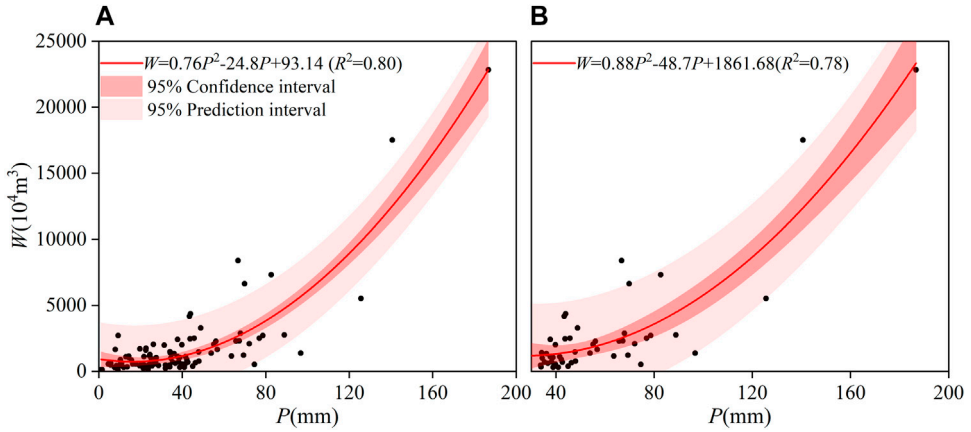


FIGURE 7 | Method of sub-rainfall fitting: **(A)** for all rainfall; **(B)** for extreme rainfall events.

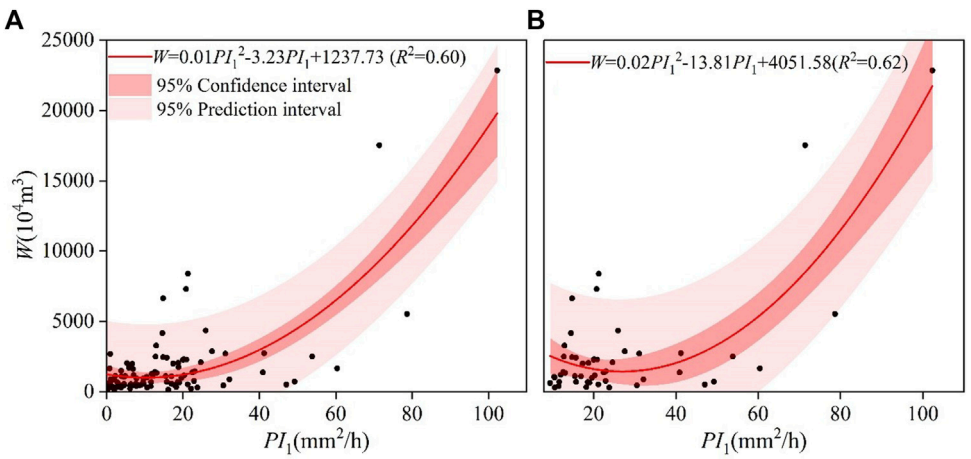


FIGURE 8 | Method of rainfall factor combination fitting: **(A)** for all rainfall; **(B)** for extreme rainfall events.

variance contribution of the seven factors is calculated. Among all rainfall events, only P_{p-6} is selected to have a significant impact on the prediction object (**Eq. 2**). For extreme rainfall events, the most significant factors affecting the prediction object are P_{P-6} and P_{2-1} (**Eq. 3**). Indicating whether it is ordinary rainfall or extreme rainfall, the accumulated rainfall (P_{p-6}) over a long period contributes significantly to runoff generation. But compared with ordinary precipitation events, the runoff yield in extreme precipitation events is gradually affected by several periods with large rainfall intensity.

$$W = 93.21P_{P-6} - 315.15 \ (R^2 = 0.60). \quad (2)$$

$$W = 117.62P_{P-6} + 665.88P_{2-1} - 4170.81 \ (R^2 = 0.69). \quad (3)$$

3.3.4 The Method of Upper Envelope

The point data on the upper edge of the rainfall–runoff figure were selected to plot the rainfall–runoff upper envelop line. The rainfall of the upper envelop points on 31 August 1995 and 3 August 1996 were 186.71 and 140.65 mm, respectively, and belonged to the type of heavy rainfall with a long duration; on 28 July 1982 and 7 September 1985, they were 82.69 and 69.41 mm, respectively, and were of the concentrated and short duration type. In both types, the upper envelop method performs better (**Figure 9**).

3.4 Extreme Rainfall Event Scenario Setting and Possible Incoming Flood Prediction

3.4.1 Extreme Rainfall Event Scenario Setting

3.4.1.1 Hydrological Frequency Analysis Method

To explore the design of extreme rainfall under different durations and different return periods, using normal, gamma, gen gamma, log gamma, generalized extreme value, Gumbel max, and Weibull distribution functions fit the possible distribution of C1 day, C2 day, C3 day, C4 day, and C5 day in the Jingle sub-basin. The optimal distribution and the corresponding parameters estimated by maximum likelihood

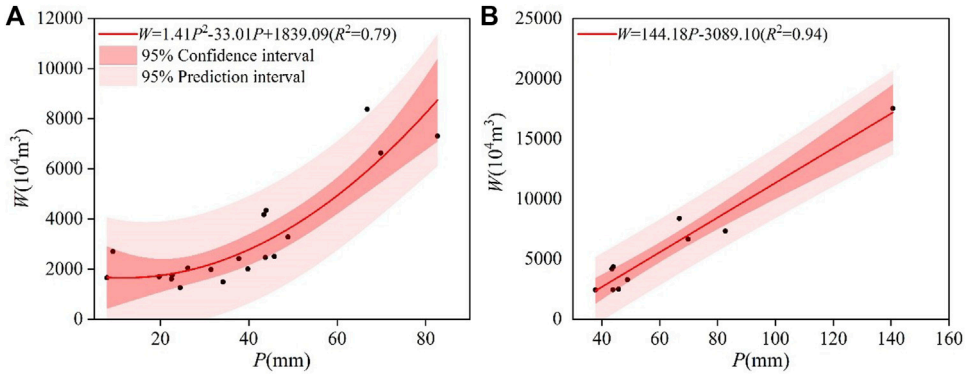


FIGURE 9 | Method of upper envelop: **(A)** for all rainfall; **(B)** for extreme rainfall events.

TABLE 3 | Optimal distribution of C1 day, C2 day, C3 day, C4 day, and C5 day.

Series	Optimal distribution	Parameter	K-S	0.05 significance level
C1 day	Normal	σ = 5.29 μ = 26.44	0.09	1.73
C2 day	Gen Gamma (4P)	k = 0.63 α = 11.38 β = 0.38 γ = 15.09	0.07	1.73
C3 day	Log Gamma	α = 219.1 β = 0.016	0.06	1.73
C4 day	Gen Gamma	k = 1.01 α = 15.75 β = 2.82	0.05	1.73
C5 day	Normal	σ = 11.47 μ = 47.00	0.06	1.73

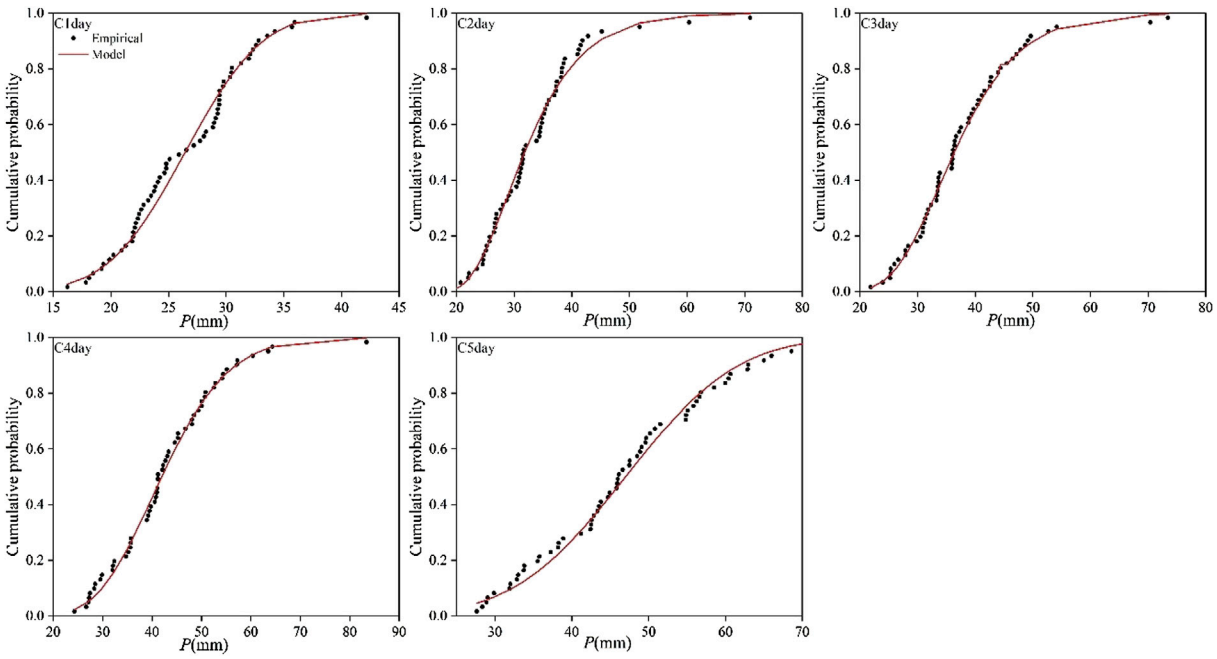


FIGURE 10 | Optimal distribution of C1 day, C2 day, C3 day, C4 day, and C5 day.

are shown in **Table 3**. **Table 3** illustrates that the normal, gen gamma (4P), log gamma, gen gamma, and normal distribution were selected as the best fitting distribution of the C1 day, C2 day, C3 day, C4 day, and C5 day, respectively (**Figure 10**). According to the parameter values of the best fitting distribution provided in **Table 3**, the value of each design extreme rainfall corresponding to different return periods can be calculated according to needs (**Table 3**). **Figure 11** shows the rainfall

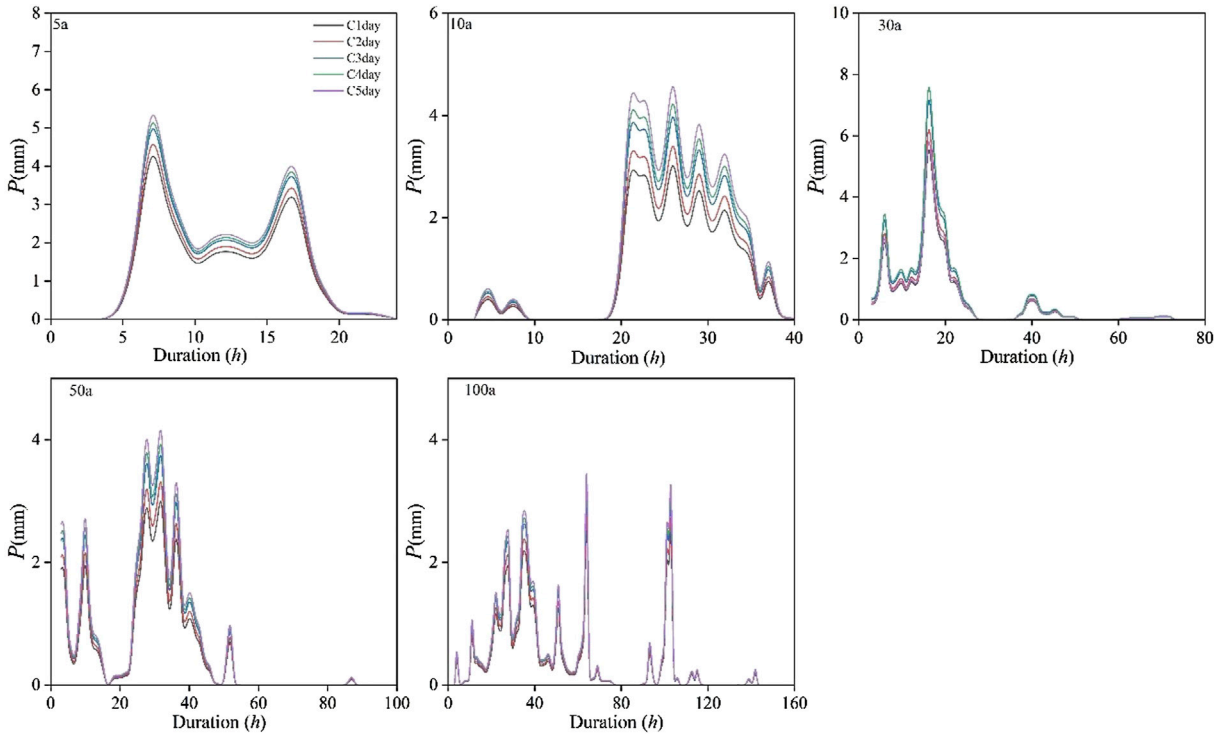


FIGURE 11 | Rainfall process under different duration and different return periods.

TABLE 4 | Extreme rainfall obtained by the method of historical extreme rainfall combination.

Date	Rainfall and rainfall in different periods (mm)									
	P	P_{p-6}	P_{6-5}	P_{5-4}	P_{4-3}	P_{3-2}	P_{2-1}	P_1	Duration	PI_1
1995/8/31	186.71	151.28	5.56	5.27	5.62	5.54	5.77	7.67	204.00	170.89
1995/8/31	186.71	151.28	5.56	5.27	5.62	5.54	5.77	7.67	204.00	170.89
2010/8/10	74.59	36.72	5.72	5.41	3.61	7.41	6.90	8.83	46.00	120.94
1973/8/19	88.89	60.55	3.76	5.57	3.61	4.75	4.16	6.49	51.00	154.95
2003/7/29	76.95	39.50	5.09	5.38	7.02	4.79	5.38	9.79	55.00	107.67
1977/7/6	43.96	5.72	3.11	5.53	6.03	8.13	7.17	8.27	18.00	107.37
2006/6/28	40.98	5.72	4.59	4.50	4.59	5.55	7.90	8.13	24.00	69.99
1974/7/22	38.26	5.24	0.42	1.18	3.60	4.95	4.85	18.01	29.00	50.47
Combined extreme rainfall	203.64	151.28	5.72	5.57	7.02	8.13	7.90	18.01	204.00	203.28

process under different duration and different return periods obtained by the fragment method.

3.4.1.2 The Method of Historical Measured Extreme Rainfall Combination

The extreme values of rainfall in each period corresponding to rainfall events are shown in **Table 4**. The extreme rainfall in each period was transplanted to the historical maximum rainfall events to obtain the rainfall process under extreme rainfall scenarios.

3.4.2 Prediction of Possible Incoming Flood Amount Under Extreme Rainfall Scenarios

The possible incoming flood amount under two extreme rainfall scenarios calculated by established rainfall–runoff empirical formulas is shown in **Table 5**. Under the same rain pattern, the predicted runoff increased with the increase in the return period. The extreme precipitation of 1–5 consecutive days under the 100a return period designed by the hydrological frequency method is 38.74, 60.01, 66.00, 71.44, and 73.69 mm, respectively, and the possible runoff predicted by the four empirical formulas is 1295–2495, 2108–4935, 2408–5801, and 3051–7062 $\times 10^4$ m^3, respectively. The rainfall designed by the hydrological frequency combination method is 203.64 mm, and the possible runoff predicted by the four runoff prediction calculation formulas is 2.8–5.3 $\times 10^8$ m^3. The runoff predicted by the method of sub-rainfall and rainfall factor combination fitting is close and less than that predicted by the method of sub-rainfall time-segment rainfall fitting and upper envelop. The method of sub-rainfall

TABLE 5 | Possible incoming flood amount under two extreme rainfall scenarios.

Scenario	Series	Return period	Design Rainfall (mm)	Rainfall characteristics of typical rainfall				Possible runoff (10^4 m^3)			
		(a)		Date	Rainfall (mm)	Runoff (10^4 m^3)	Duration (h)	Formula 1	Formula 2	Formula 3	Formula 4
S1	C1 day	5a	30.88	1977/7/22	37.77	1102.98	21	1197.00	987.74	856.41	1363.90
		10a	33.21	1977/7/22	37.77	1102.98	21	1214.92	999.52	1073.22	1699.27
		30a	36.13	1977/7/22	37.77	1102.98	21	1250.91	1020.39	1345.45	2120.36
		50a	37.29	1977/7/22	37.77	1102.98	21	1269.40	1030.57	1453.79	2287.95
		100a	38.74	1977/7/22	37.77	1102.98	21	1295.64	1044.69	1588.14	2495.76
	C2 day	5a	39.62	1988/7/17	56.10	1049.62	42	1313.60	2435.88	675.59	2744.73
		10a	44.69	1988/7/17	56.10	1049.62	42	1442.93	2293.40	1272.16	3180.29
		30a	52.18	1988/7/17	56.10	1049.62	42	1716.32	2109.89	2152.07	3955.14
		50a	55.53	1988/7/17	56.10	1049.62	42	1870.87	2037.92	2546.57	4353.77
		100a	60.01	1988/7/17	56.10	1049.62	42	2108.02	1951.82	3073.15	4935.31
	C3 day	5a	44.82	1994/7/6	47.84	1459.98	54	1446.84	1704.69	1687.77	3192.41
		10a	50.14	1994/7/6	47.84	1459.98	54	1632.12	1668.67	2183.08	3728.52
		30a	57.89	1994/7/6	47.84	1459.98	54	1991.54	1708.21	2905.65	4653.41
		50a	61.36	1994/7/6	47.84	1459.98	54	2186.57	1761.27	3228.90	5122.04
		100a	66.00	1994/7/6	47.84	1459.98	54	2480.56	1866.45	3661.30	5801.93
	C4 day	5a	51.47	1976/8/19	67.30	762.68	91	1686.39	1935.80	2040.87	3875.49
		10a	56.95	1976/8/19	67.30	762.68	91	1942.42	1834.57	2685.54	4532.48
		30a	64.35	1976/8/19	67.30	762.68	91	2371.78	1735.81	3555.58	5553.45
		50a	67.46	1976/8/19	67.30	762.68	91	2580.78	1707.31	3920.91	6028.15
		100a	71.44	1976/8/19	67.30	762.68	91	2873.63	1681.98	4389.39	6676.70
	C5 day	5a	56.66	1977/6/23	72.13	2087.32	99	1927.22	1777.98	1786.84	4494.76
		10a	61.70	1977/6/23	72.13	2087.32	99	2207.04	1717.92	2380.35	5170.28
		30a	68.04	1977/6/23	72.13	2087.32	99	2621.83	1675.51	3125.54	6120.13
		50a	70.56	1977/6/23	72.13	2087.32	99	2806.56	1668.84	3422.12	6529.64
		100a	73.69	1977/6/23	72.13	2087.32	99	3051.18	1668.66	3789.88	7062.35
S2		Combined rainfall	203.64		--		204	28,437.67	2070.72	18,886.46	53,589.13

S1 is the scenario obtained by the hydrological frequency analysis method; S2 is the scenario obtained by the historical measured extreme rainfall combination method.

time-segment rainfall fitting and the upper envelop method are more sensitive to the change of rainfall and increase with the return period.

In this study, based on the daily precipitation data of the past 6 decades, the study of the 95th percentile threshold of daily to 5 consecutive days of precipitation shows that the trend of extreme precipitation presents a no significant increase. Gao et al., 2015 and Sun et al., 2020 obtained the same result by studying other rainfall indicators. The characteristics of rainfall events such as precipitation, rainfall intensity, and rainfall duration have been

demonstrated, which is closely related to the amount of flood (Ran et al., 2012; Adib et al., 2018). For all rainfall events and selected extreme rainfall events, different rainfall factors were selected to establish several runoff prediction formulas, which directly map the relationship between rainfall and runoff and objectively respond to the influence of a variable on water yield. The accuracy of runoff prediction mainly depends on the correlation between selected rainfall factors and runoff (Quilty et al., 2016). Rainfall is the direct source of runoff; it has a high correlation with runoff. There was the closest relationship between rainfall and runoff. Many researchers established empirical hydrological models based on the relationship between rainfall and runoff by using regression analysis, which achieved desirable results (Sharifi et al., 2017; Zeinali et al., 2019). In this study, the correlation coefficient between rainfall and runoff reached 0.85, and the fitting effect was relatively satisfactory (**Figure 7**). In semi-arid areas, rainfall intensity is particularly important for surface runoff generation (Smith et al., 2010; Yang et al., 2016). The combination of the rainfall factors method integrates the effect of rainfall amount and the maximum 1 h hour intensity on water and sediment yield in the watershed and it has a better fitting effect on the two rainfall types of large amounts, long duration, and high concentration and intensity (**Figure 8**). This is consistent with the research findings of Huang et al., 2010. The water and sediment yield of most rainstorms in The Loess Plateau Basin is concentrated in a few periods (Zhou and Wang, 1992). In this study, stepwise regression shows whether it is ordinary rainfall or extreme rainfall, and the accumulated rainfall (P_{p-6}) over a long period contributes greatly to runoff generation. But, compared with ordinary rainfall events, the runoff yield in extreme precipitation events is gradually affected by several periods with large rainfall intensity. This phenomenon is more obvious in a rainstorm (Li et al., 2016). The upper envelop method considers the extreme cases under the rainfall–runoff event. (**Figure 9**). The four formulas all show that there is a more significant correlation between rainfall factors and runoff under extreme rainfall events. The relationship between rainfall characteristics and runoff generation increases with the increase in rainfall level and intensity (Zhang et al., 2020; Jonathan et al., 2021). Although the characteristics of a rainfall event are closely related to the amount of flood, it seems that this relationship is more obvious in extreme rainfall events. So, it is necessary to study the relationship between extreme rainfall events and runoff alone, and it is of great significance for the sustainable utilization of water resources and flood control and disaster reduction in the basin.

4 CONCLUSION

Based on the daily precipitation data of the past 6 decades, the trend characteristics of the 95th percentile extreme precipitation (daily to 5 consecutive days of precipitation) were studied. Moreover, different rainfall factors were selected to establish several runoff prediction formulas for all rainfall events and selected extreme rainfall events and then the empirical formulas were combined with designed extreme rainfall scenarios to predict the possible incoming flood. It can provide a reference for water security and major project deployment in the basin.

The main findings of the present study are summarized as follows: the 95th percentile extreme precipitation (daily to 5 consecutive days of precipitation) of the Jingle sub-basin presents a slight increased trend and has positive consistency. Combining the results of MK and R/S, the extreme precipitation in this basin will maintain a slight rising trend in the future, assuming that the climate and underlying conditions remain the same as they were in the current scenario. There is a more stable correlational relationship between rainfall factors (*P*, *PI*, seven sub-periods) and runoff in extreme rainfall events. The extreme precipitation of 1–5 consecutive days under the 100a return period designed by the hydrological frequency method is 38.74, 60.01, 66.00, 71.44, and 73.69 mm, respectively, and the possible runoff predicted by the four empirical formulas is 1295–2495, 2108–4935, 2408–5801, and 3051–7062 $\times 10^4$ m^3, respectively. The rainfall designed by the hydrological frequency combination method is 203.64 mm, and the possible runoff predicted by the four empirical formulas is 2.8–5.3 $\times 10^8$ m^3. Although the characteristics of rainfall events are closely related to the amount of flood, it seems that this relationship is more obvious in extreme rainfall events. So, it is necessary to study the relationship between extreme rainfall events and runoff alone, and it is of great significance for the sustainable utilization of water resources and flood control and disaster reduction in the basin.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

SJ development or design of methodology and creation of models. CY application of statistical, mathematical, computational, or other formal techniques to analyze or synthesize study data. YW verification. XY and YL acquisition of data.

FUNDING

This project was supported by the Natural Science Foundation of Henan Province (212300410413), the Young Talents Promotion Project of Henan Province (2021HYTP030), and the China Postdoctoral Science Foundation (2020M672247); State Key Laboratory of Soil Erosion and Dryland Farming on the Loess Plateau Foundation (A314021402-2021011).

REFERENCES

Adib, A., Lotfirad, M., and Haghighi, A. J. (2018). Using Uncertainty and Sensitivity Analysis for Finding the Best Rainfall-Runoff Model in Mountainous Watersheds (Case Study: the Navrood Watershed in Iran). *J. Mt. Sci.* 16 (3), 529–543. doi:10.1007/s11629-018-5010-6

Barlow, M., Gutowski, W. J., Gyakum, J. R., Katz, R. W., Lim, Y.-K., Schumacher, R. S., et al. (2019). North American Extreme Precipitation Events and Related Large-Scale Meteorological Patterns: a Review of Statistical Methods, Dynamics, Modeling, and Trends. *Clim. Dyn.* 53 (11), 6835–6875. doi:10.1007/s00382-019-04958-z

Dang, S., Liu, X., Yin, H., and Guo, X. (2020). Prediction of Sediment Yield in the Middle Reaches of the Yellow River Basin under Extreme Precipitation. *Front. Earth Sci.* 8, 542686–542699. doi:10.3389/feart.2020.542686

de Lima, J. L. M. P., Tavares, P., Singh, V. P., and de Lima, M. I. P. (2009). Investigating the Nonlinear Response of Soil Loss to Storm Direction Using a Circular Soil Flume. *Geoderma* 152 (1–2), 9–15. doi:10.1016/j.geoderma.2009.05.004

Galelli, S., and Castelletti, A. (2013). Assessing the Predictive Capability of Randomized Tree-Based Ensembles in Streamflow Modelling. *Hydrol. Earth Syst. Sci.* 17 (7), 2669–2684. doi:10.5194/hess-17-2669-2013

Gao, Y., Feng, Q., Liu, W., Lu, A. G., Wang, Y., and Yang, J. (2015). Changes of Daily Climate Extremes in Loess Plateau during 1960–2013. *Quat. Int.* 371, 5–21. doi:10.1016/j.quaint.2014.08.052

Gökbulak, F., Şengönül, K., Serengil, Y., Yurtseven, İ., Özhan, S., Cigizoglu, H. K., et al. (2015). Comparison of Rainfall-Runoff Relationship Modeling Using Different Methods in a Forested Watershed. *Water Resour. Manag.* 29 (12), 4229–4239. doi:10.1007/s11269-015-1056-1

Hao, G., Li, J., Song, L., Li, H., and Li, Z. (2018). Comparison between the TOPMODEL and the Xin'anjiang Model and Their Application to Rainfall Runoff Simulation in Semi-humid Regions. *Environ. Earth Sci.* 77 (7), 279. doi:10.1007/s12665-018-7477-4

Huang, Z., OuYang, Z., Li, F., Zheng, H., and Wang, X. (2010). Response of Runoff and Soil Loss to Reforestation and Rainfall Type in Red Soil Region of Southern China. *J. Environ. Sci.* 22 (11), 1765–1773. doi:10.1016/s1001-0742(09)60317-x

Jonathan, F., Frederik, K., Daniel, K., Martin, G., Guy, S., and Oren, G. (2021). Deep Learning Rainfall-Runoff Predictions of Extreme Events. *Hydrology Earth Syst. Sci.* 91, 179–199. doi:10.5194/hess-2021-423

Li, F., Tang, G., Wang, C., Cui, L., and Zhu, R. (2016). Slope Spectrum Variation in a Simulated Loess Watershed. *Front. Earth Sci.* 10 (2), 328–339. doi:10.1007/s11707-015-0519-2

Li, W., Yang, M., Liang, Z., Zhu, Y., Mao, W., Shi, J., et al. (2013). Assessment for Surface Water Quality in Lake Taihu Tiaoxi River Basin China Based on Support Vector Machine. *Stoch. Environ. Res. Risk Assess.* 27 (8), 1861–1870. doi:10.1007/s00477-013-0720-3

Liu, X. Y., Dang, S. Z., and Gao, Y. F. (2019). Sediment Yield of Current Underlying Surface under Simulated Extreme Rainstorm in Middle Reaches of Yellow River Basin. *Trans. Chin. Soc. Agric. Eng.* 35 (11), 131–138. doi:10.11975/j.issn.1002-6819.2019.11.015

Manfreda, S., Iacobellis, V., Gioia, A., Fiorentino, M., and Kochanek, K. (2018). The Impact of Climate on Hydrological Extremes. *Water* 10 (6), 802–805. doi:10.3390/w10060802

Qamar, M. U., Azmat, M., Cheema, M. J. M., Shahid, M. A., Khushnood, R. A., and Ahmad, S. (2016). Model Swapping: A Comparative Performance Signature for the Prediction of Flow Duration Curves in Ungauged Basins. *J. Hydrology* 541, 1030–1041. doi:10.1016/j.jhydrol.2016.08.012

Qi, L., Wang, D. W., Zhang, Y. L., and Wang, L. (2021). Flood Simulation Analysis of the Biliu River Basin Based on the MIKE Model. *Complexity* 10. doi:10.1155/2021/8827046

Quilty, J., Adamowski, J., Khalil, B., and Rathinasamy, M. (2016). Bootstrap Rank-Ordered Conditional Mutual Information (broCMI): A Nonlinear Input Variable Selection Method for Water Resources Modeling. *Water Resour. Res.* 52 (3), 2299–2326. doi:10.1002/2015wr016959

Ran, Q., Su, D., Li, P., and He, Z. (2012). Experimental Study of the Impact of Rainfall Characteristics on Runoff Generation and Soil Erosion. *J. Hydrology* 424-425 (6), 99–111. doi:10.1016/j.jhydrol.2011.12.035

Seckin, N., Cobaner, M., Yurtal, R., and Haktanir, T. (2013). Comparison of Artificial Neural Network Methods with L-Moments for Estimating Flood Flow at Ungauged Sites: the Case of East Mediterranean River Basin, Turkey. *Water Resour. Manage* 27, 2103–2124. doi:10.1007/s11269-013-0278-3

Sharifi, A., Dinpashoh, Y., and Mirabbasi, R. (2017). Daily Runoff Prediction Using the Linear and Non-linear Models. *Water Sci. Technol.* 76 (4), 793–805. doi:10.2166/wst.2017.234

Shi, J., Cui, L., Wen, K., Tian, Z., Wei, P., and Zhang, B. (2018). Trends in the Consecutive Days of Temperature and Precipitation Extremes in China during 1961-2015. *Environ. Res.* 161, 381–391. doi:10.1016/j.envres.2017.11.037

Shi, J., Wen, K., and Cui, L. (2016b). Patterns and Trends of High-Impact Weather in China during 1959-2014. *Nat. Hazards Earth Syst. Sci.* 16, 855–869. doi:10.5194/nhess-16-855-2016

Smith, M. W., Bracken, L. J., and Cox, N. J. (2010). Toward a Dynamic Representation of Hydrological Connectivity at the Hillslope Scale in Semiarid Areas. *Water Resour. Res.* 46 (12), 65–74. doi:10.1029/2009wr008496

Sun, C., Huang, G., and Fan, Y. (2020). Multi-Indicator Evaluation for Extreme Precipitation Events in the Past 60 Years over the Loess Plateau. *Water* 12 (1), 193. doi:10.3390/w12010193

Sun, Q., Miao, C., Duan, Q., and Wang, Y. (2015). Temperature and Precipitation Changes over the Loess Plateau between 1961 and 2011, Based on High-Density Gauge Observations. *Glob. Planet. Change* 132, 1–10. doi:10.1016/j.gloplacha.2015.05.011

Tan, X., Wu, X., and Liu, B. (2021). Global Changes in the Spatial Extents of Precipitation Extremes. *Environ. Res. Lett.* 16 (5), 054017. doi:10.1088/1748-9326/abf462

Teng, F., Huang, W., and Ginis, I. (2017). Hydrological Modeling of Storm Runoff and Snowmelt in Taunton River Basin by Applications of HEC-HMS and PRMS Models. *Nat. Hazards* 91, 179–199. doi:10.1007/s11069-017-3121-y

Visessri, S., and McIntyre, N. (2016). Regionalisation of Hydrological Responses under Land-Use Change and Variable Data Quality. *Hydrological Sci. J.* 61 (2), 302–320. doi:10.1080/02626667.2015.1006226

Wang, Y., Guo, S., Chen, H., and Zhou, Y. (2014). Comparative Study of Monthly Inflow Prediction Methods for the Three Gorges Reservoir. *Stoch. Environ. Res. Risk Assess.* 28 (3), 555–570. doi:10.1007/s00477-013-0772-4

Westra, S., Alexander, L. V., and Zwiers, F. W. (2013). Global Increasing Trends in Annual Maximum Daily Precipitation. *J. Clim.* 26 (11), 3904–3918. doi:10.1175/jcli-d-12-00502.1

Wu, J., Zheng, H., and Xi, Y. (2019). SWAT-based Runoff Simulation and Runoff Responses to Climate Change in the Headwaters of the Yellow River, China. *Atmosphere* 10, 509. doi:10.3390/atmos10090509

Wu, S., Hu, Z., Wang, Z., Cao, S., Yang, Y., Qu, X., et al. (2021). Spatiotemporal Variations in Extreme Precipitation on the Middle and Lower Reaches of the Yangtze River Basin (1970-2018). *Quat. Int.* 592, 80–96. doi:10.1016/j.quaint.2021.04.010

Yang, T., Wang, Q., Su, L., Wu, L., Zhao, G., Liu, Y., et al. (2016). An Approximately Semi-analytical Model for Describing Surface Runoff of Rainwater over Sloped Land. *Water Resour. Manage* 30 (11), 3935–3948. doi:10.1007/s11269-016-1400-0

Zeinali, V., Vafakhah, M., and Sadeghi, S. H. (2019). 'Impact of Urbanization on Temporal Distribution Pattern of Storm Runoff Coefficient. *Environ. Monit. Assess.* 191 (9), 595. doi:10.1007/s10661-019-7734-3

Zhang, J. P., Zhang, H., Xiao, H. L., and Fang, H. Y. (2020). Effects of Rainfall and Runoff-Yield Conditions on Runoff. *Ain Shams Eng. J.* 12, 2111–2116. doi:10.1016/j.asej.2020.10.010

Zhang, Y., Chiew, F. H. S., Li, M., and Post, D. (2018). Predicting Runoff Signatures Using Regression and Hydrological Modeling Approaches. *Water Resour. Res.* 54 (10), 7859–7878. doi:10.1029/2018wr023325

Zhao, Y., Liu, B., Zhang, X. M., and Bao, S. Z. (2020). Study on Flood and Sediment Yield under Extreme Precipitation in the Wuding River Basin of the Yellow

River. *J. Sediment Res.* 45 (06), 47–52. doi:10.16239/j.cnki.0468-155x.2020.06.008

Zhou, P. H., and Wang, Z. L. (1992). A Study on Rainstorm Causing Soil Erosion in the Loess Plateau. *J. Soil Water Conservation* 6 (3), 1–5.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer HW declared a shared affiliation with the authors JS, YC, and WY to the handling editor at the time of review

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Environmental Science

ORIGINAL RESEARCH
published: 01 July 2022
doi: 10.3389/fenvs.2022.887323

Projection of Precipitation Extremes and Flood Risk in the China–Pakistan Economic Corridor

Shixiong Du[1], Ruiying Wu[1,2], Huaiwei Sun[1], Dong Yan[1], Jie Xue[3]*, Weihong Liao[4], Ye Tuo[5] and Wenxin Zhang[6]*

[1]School of Civil and Hydraulic Engineering, Huazhong University of Science and Technology, Wuhan, China, [2]Water Conservancy Information and Propaganda Education Center, Yancheng, China, [3]State Key Laboratory of Desert and Oasis Ecology, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi, China, [4]China Institute of Water Resources and Hydropower Research, Beijing, China, [5]Chair of Hydrology and River Basin Management, Technical University of Munich, Munich, Germany, [6]Department of Physical Geography and Ecosystem Science, Lund University, Lund, Sweden

It is reported that the China–Pakistan Economic Corridor has been affected by extreme precipitation events. Since the 20th century, extreme weather events have occurred frequently, and the damage and loss caused by them have increased. In particular, the flood disaster caused by excessive extreme precipitation seriously hindered the development of the human society. Based on CRiteria Importance Through Intercriteria Correlation and square root of generalized cross-validation, this study used intensity–area–duration to analyze the trend of future extreme precipitation events, corrected the equidistance cumulative distribution function method deviation of different future scenario models (CESM2, CNRM-CM6-1, IPSL-CM6A-LR, and MIROC6) and evaluated the simulation ability of the revised model. The results showed that: 1) the deviation correction results of CNRM-CM6-1 in the Coupled Model Intercomparison Project Phase (CMIP) 6 could better simulate the precipitation data in the study area, and its single result could achieve the fitting effect of the CMIP5 multimodel ensemble average; 2) under CNRM-CM6-1, the frequency of extreme precipitation events under the three climate scenarios (SSP1-2.6, SSP3-7.0, and SSP5-8.5) presents interdecadal fluctuations of 3.215 times/10A, 1.215 times/10A, and 5.063 times/10A, respectively. The average impact area of extreme precipitation events would decrease in the next 30 years, while the total impact area and the extreme precipitation events in a small range would increase. Under the future scenario, the increase rate of extreme precipitation was highest in August, which increased the probability of extreme events; 3) in the next 30 years, the flood risk had an obvious expansion trend, which was mainly reflected in the expansion of the area of high-, medium-, and low-risk areas. The risk zoning results obtained by the two different flood risk assessment methods were different, but the overall risk trend was the same. This study provided more advanced research for regional flood risk, reasonable prediction for flood risk under future climate models, and useful information for flood disaster prediction in the study area and contributes to the formulation of local disaster prevention and reduction policies.

Keywords: China–Pakistan economic corridor, precipitation extremes, IAD, future scenarios, flood risk

OPEN ACCESS

Edited by:
Fei Tian,
China Agricultural University, China

Reviewed by:
Sher Muhammad,
International Centre for Integrated Mountain Development, Nepal
Safi Ullah,
Fudan University, China

****Correspondence:***
Huaiwei Sun
hsun@hust.edu.cn
Jie Xue
xuejie11@mails.ucas.ac.cn

Specialty section:
This article was submitted to Environmental Economics and Management, a section of the journal Frontiers in Environmental Science

Received: *01 March 2022*
Accepted: *24 May 2022*
Published: *01 July 2022*

Citation:
Du S, Wu R, Sun H, Yan D, Xue J, Liao W, Tuo Y and Zhang W (2022) Projection of Precipitation Extremes and Flood Risk in the China–Pakistan Economic Corridor. Front. Environ. Sci. 10:887323. doi: 10.3389/fenvs.2022.887323

1 INTRODUCTION

The impact of climate change on the hydrological cycle has been recognized for a long time (Sun et al., 2021). Although it is widely acknowledged that precipitation extremes are likely to cause an increase in flood risk, the relationship between climate and flood is rather complex (Zhang et al., 2008). It was reported that frequent extreme climate events had brought the serious loss of life and property to people worldwide since the beginning of the 21st century (Zhang et al., 2011). With the increasing emissions of global carbon dioxide and other greenhouse gases, global warming continues to intensify (Pachauri and Reisinger, 2008; Huang et al., 2017), and the instability and extremes of the climate increased, increasing the intensity and frequency of extreme precipitation events and floods in the future scenario (Meehl et al., 2000). Floods caused by extreme precipitation frequently occur in China-Pakistan Economic Corridor (CPEC) and increased frequency and strength with duration expansion. Therefore, it was urgently needed to investigate climate change's flood risk in high-risk areas.

Extreme precipitation events are critical indicators for studying extreme climate events and an essential factor for studying future climate changes. The flood disaster caused by extreme precipitation seriously hinders the development of society, and human civilization's progress has become the focus of attention all over the world (Goswami et al., 2006). In the present-day climate over most of the globe, the curve relating daily precipitation extremes with local temperatures had a peak structure, increasing as expected at the low–medium range of temperature variations but decreasing at high temperatures (Wang et al., 2011; Chang et al., 2022). Wang et al. (2015) attempted to explain the climate change effects on regional precipitation. However, the characteristics of precipitation extremes may depend on the method used for analysis. One method that could reveal precipitation characteristics is Intensity-Area-Duration (IAD), which for identifying extreme precipitation events was improved based on severity–Area–Duration (SAD) of Andreadis et al. (2005). This method was proved to be effective for assessing drought and flood risk by several researchers. Jing et al. (2016) applied IAD to identify regional extreme precipitation events for the first time in researching regional extreme precipitation events in China. They correlated the identified extreme precipitation events with population economic exposure. Wen et al. (2019) also used this method to identify drought events under three global warming scenarios based on several global climate models (CPEC (formerly known as silk road and well-known as Karakoram Highway) had been affected by extreme 25 precipitation events).

It should be noted that the role of precipitation extremes in shaping flood risk depends on land cover, region, and environmental conditions (Wang et al., 2017; Sun et al., 2022). Benito et al. (2015) used the flood risk assessment model established by the ancient flood data to assess the risk in Europe. The ancient flood data had a large time span and can fully reflect the impact of climate change. This kind of assessment method was more accurate in calculating flood probability. Yang et al. (2010) used the BP neural network algorithm of rough set reduction to obtain the flood risk. The integrated system based on the spatial processing ability of the geographic information system (GIS) has gradually become a powerful tool for flood risk assessment. Brendel et al. (2021) used SWMM and GSSHA to model storm pipeline networks and urban floods in Roanoke, Virginia. They found that the value of GSSHA to the city lies in its ability to predict flood duration and spatial range in a two-dimensional rangeability.

The GCMs were considered to be useful for the investigation of hydrological cycles (Li et al., 2018; Yang et al., 2021), decision-making in water resource management (Sun et al., 2021), and the atmosphere–land interactions (Simpkins, 2017; Sun et al., 2021). It may be helpful to use future scenarios and GCMs for the projection of precipitation extremes and flood risks (Su et al., 2008). The climate model was a vital tool used to predict climate change and explore the change mechanism of meteorological elements (Xue et al., 2013). At present, GCM simulations have provided climate change scenarios for scholars worldwide to carry out future climate research and evaluation and climate negotiations (Zhao et al., 2021). Regionalized increased and decreased drought duration and frequency were driven by changes in precipitation mean and variability (Su et al., 2006; Pierce et al., 2009; Sun et al., 2016). To predict the future flood risk of CPEC, provide theoretical support for managers to formulate policies, and reduce the losses caused by extreme events in CPEC, this study used IAD to predict the trend of future extreme precipitation events based on the data on three new combined path models in CMIP6. In addition, it used four future scenario model data sets to predict future extreme events. The results were analyzed and compared by downscaling analysis. Then, the most suitable model for CPEC was selected to obtain the development trend of flood disaster risk of CPEC and provide theoretical support for future risk aversion. The rest of this article describes the data and methods in **Section 2**. The results are presented in **Section 3**, followed by discussions in **Section 4** and conclusions in **Section 5**. This study will provide a reference for the research of regional flood risk and the prediction of regional flood risk under future scenarios and provide theoretical support for extreme regional events and flood prevention measures.

2 DATA AND METHODS

2.1 Study Area

CPEC extends from the port of Guarda in Pakistan to Kashgar in China, especially covering the whole territory of Pakistan, Kashgar in Xinjiang, and its surrounding areas, with a total length of about 3,000 km and a total area of about 932,000 km^2. The Indus River is an international river that runs through the whole territory of Pakistan and provides most of the irrigation water in the region. Its five tributaries, Jhelum River, Janab River, Ravi River, Bias River, and Sutlej River, converge in the Punjab plain (**Figure 1**). The precipitation in CPEC was mainly affected by two weather systems: summer precipitation and winter precipitation (Khan et al., 2014). Summer precipitation resulted from the Indian Ocean

TABLE 1 | Basic information of four global climate models in CMIP6.

Model	Institution	Resolution (longitude × latitude)	Ensembles
CESM2	NSF-DOE-NCAR	1.25 × 0.9424	r1i1p1f1
CNRM-CM6-1	CNRM-CERFACS	1.4062 × 1.4088	r1i1p1f2, r2i1p1f2, r3i1p1f2, r4i1p1f2, r5i1p1f2, and r6i1p1f2
IPSL-CM6A-LR	IPSL	2.5 × 1.2676	r1i1p1f1
MIROC6	MIROL	1.4062 × 1.4088	r1i1p1f1

monsoon disturbance, and winter precipitation resulted from the Mediterranean westerly disturbance (Safi et al., 2018). The southern part of CPEC was affected by the Indian Ocean monsoon climate with uneven precipitation and regional precipitation within the year, which was very prone to extreme precipitation. Since the 1990s, the precipitation in this area has increased significantly, and extreme climate events have been significant. Flash floods, originating from extreme weather events, have relatively less duration but severe intensity and impacts. These floods usually occur during the South Asian monsoon period between July and September (Memon et al., 2015). In 2011, large-scale heavy rains were observed in Sindh, leading to substantial economic losses, destruction of ecological resources, food shortages, and starvation (Haq et al., 2012).

Historically, CPEC has suffered many rainstorms and flood disasters (Federal Flood Commission, Ministry of Water and Power, 2015). According to statistics, 25 significant flood events have occurred in CPEC in the past 70 years. The flood disaster had caused more than $30 billion in the loss in Pakistan. About 25,502 people were killed, 197,273 villages were destroyed, 616,598 km^2 of land has been affected, and the flood disaster has become one of the main challenges affecting local economic and social development.

2.2 Data

2.2.1 Precipitation Data Sets

This study used a specific precipitation data set to study and analyze the extreme precipitation events and flood risk assessment. This data set used the professional meteorological interpolation software ANUSPLIN to carry out spatial interpolation combined with three-dimensional geospatial information and evaluate the interpolation model's effect through generalized cross-validation and error analysis. Test and verification of this data set can be referred to Wu et al. (2021).

The CMIP6 precipitation data were selected for this study's extreme precipitation and flood risk projection. Compared with the planned model, the scenario model in CMIP6 usually had better resolution and improved dynamic process, and the new emission scenario based on the shared socioeconomic pathway (SSP)/Representative Concentration Pathways (RCP) could be used for future climate change simulation (Eyring et al., 2016; O'Neill et al., 2016; Riahi et al., 2017; Jiang et al., 2020). The models data used in this study were three scenarios (SSP1-2.6, SSP3-7.0, and SSP5-8.5) of the global climate model (**Table 1**) in the scenario comparison plan under CMIP6. The precipitation data from 1984 to 2013 were used as the base period, and the data from 2021 to 2050 were used as the simulation data. In order to facilitate comparative analysis, based on the observation data set, the spatial resolution of model data was uniformly interpolated on a grid point of 0.25 × 0.25 by bilinear interpolation. The data used in the grid include the data on 50 meteorological stations within CPEC. The stations are evenly distributed in the south of 30°N, relatively concentrated in 30°N ~ 35°N, and almost no stations are distributed in the north of 35°N in the study area. According to the precipitation distribution in the study area, the area north of 35°N is affected by terrain and airflow, and the annual precipitation is sparse, so it can be ignored. The precipitation in the whole study area is mainly distributed in the southern and central plains. Therefore, the selection of sites is reasonable.

Notably, the daily scale data are more obvious than the monthly scale data. We had considered using daily scale data to make the results of the benchmark period close to the extreme situation. However, the data on the benchmark period here was mainly to verify the relative accuracy of the model simulation and select the best. Therefore, the data on the monthly scale could fully achieve the purpose here. We also referred that Ali et al. (2018) used Hydrologiska Byrans Vattenbalansavdeling (HBV) light model to simulate the hydrology of the Hunza River Basin, which is affected by extreme precipitation. It was found that the model based on monthly scale data performs better. Based on this, they compared the simulation results of CSM1.1, CanESM2, and MIROC-ESM three GCMs models in the future scenario.

Here, this study selected four different model data recently released by CMIP6 for long-term simulation ability evaluation and set the model data closest to the measured data for statistical downscaling to improve the simulation ability of the model further (Lu et al., 2021). These models were commonly used by others. Abbas et al. (2022) used these four models for climate simulation in Pakistan. The CESM2 simulations exhibit agreement with satellite-era observations of the climate mean state, seasonal cycle, and interannual variability that are among the closest coupled climate model in the present CMIP6 archive (Danabasoglu et al., 2020). The equilibrium climate sensitivity of CNRM-CM6-1 is significantly increased compared to that of CNRM-CM5-1 (Voldoire et al., 2019). The equilibrium climate sensitivity and transient climate response of IPSL-CM6A-LR have increased from the previous climate model IPSL-CM5A-LR used in CMIP5 (Boucher et al., 2020). The tropical climate systems (e.g., summertime precipitation in the western Pacific and the eastward-propagating Madden-Julian oscillation) and the midlatitude atmospheric circulation (e.g., the westerlies, the polar

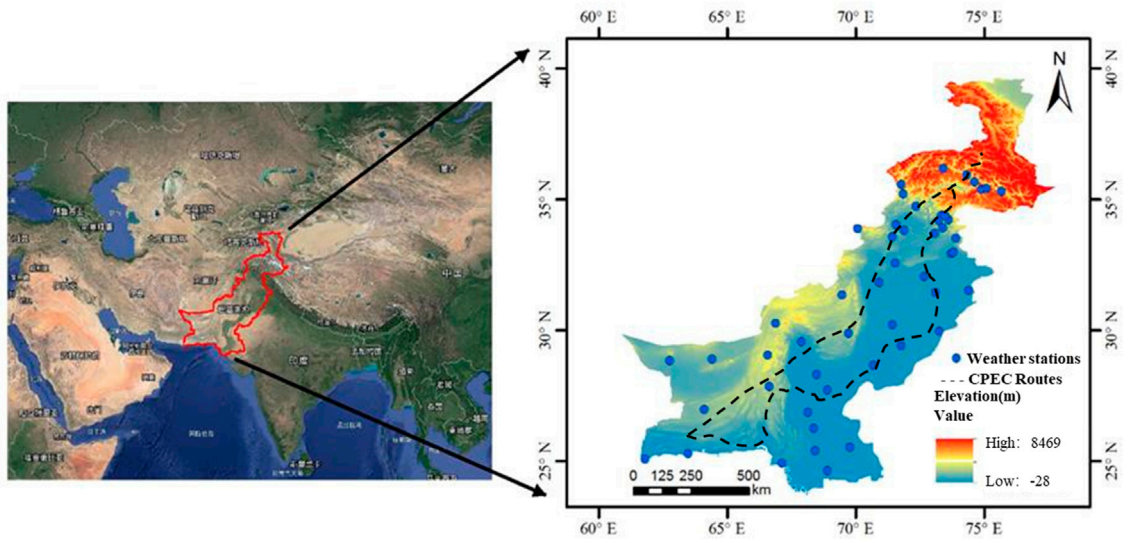


FIGURE 1 | Geographical location and meteorological station distribution of CPEC.

night jet, and troposphere–stratosphere interactions) are significantly improved in MIROC6 (Kataoka et al., 2020).

2.2.2 Collected Data for Flood Risk Assessment

The CPEC regional geographic data set was constructed in this study by comprehensively considering multiple data sources. The resampling method was used to solve the spatial data resolution difference between multiple data sources. The selected DEM data with 30 m resolution (downloaded from geospatial data cloud: http://www.gscloud.cn/) and the original DEM data processing were used to obtain non-depression DEM and slope data. The land use was divided into 10 categories (the impact of vegetation, water conservancy facilities, and other factors considered in the classification). Globeland30 (30M global surface coverage data, downloaded from the National Geographic Information Resources Directory Service System: https://www.webmap.cn/) was used to sort according to the impact degree, and the partial area index of the cultivated land was obtained after processing. The NDVI indexes required for analysis were obtained from landsat8 satellite data through image processing and band operation. Furthermore, the gridded population of the world (GPW) V4 population density data set (grid 0.25°, 30 km resolution) was selected to obtain the population density data in the study area. In order to get the building density, this study used the 2010 QuickBird orthophoto of CPEC as the primary data source. The road network density map was drawn after the aforementioned parallel processing based on the road network data (from OpenStreetMap: https://www.openstreetmap.org/).

2.3 Methods

2.3.1 Intensity–Area–Duration (IAD)

By employing IAD, this study comprehensively considered the three-dimensional characteristics, which was the intensity, impact area, and duration of extreme precipitation, and defines the grid set with certain intensity within a specific range on a given timescale as an extreme precipitation event based (Andreadis et al., 2005) on drought SAD.

In identifying extreme precipitation events, we first extracted the scope of the event, that is, the influence area of an extreme precipitation event. The grid point with the largest relative intensity within the range of extreme precipitation events in each timescale was the "strongest precipitation center" of the event, and then we searched the "secondary heavy precipitation center" from the center to the surrounding and repeated until there was no point exceeding the threshold within the range. Each extreme precipitation event's relative intensity and impact area was recorded, and then we found a new "strongest precipitation center" and repeated the aforementioned steps until all regional extreme precipitation events within the duration scale were found (**Figure 2**).

2.3.2 Assessment of the Flood Risk

The intuitionistic fuzzy analytic hierarchy process (IFAHP) was an improved subjective weighting method based on the analytic hierarchy process (AHP) (Sadiq and Tesfamariam, 2009). First, the intuitionistic fuzzy judgment matrix was constructed, and then its consistency was tested, and finally, the weight of each index was calculated.

CRITIC was an objective weighting method proposed by Diakoulaki et al. (1995). The basic idea of determining the index weight was based on two fundamental concepts: one was a comparative strength, and the other was the conflict between indicators. The basic idea of the critical method was to comprehensively use the difference and disagreement between indicators to calculate the weight, and the difference was based on the standard deviation σ. The calculation formula was as follows:

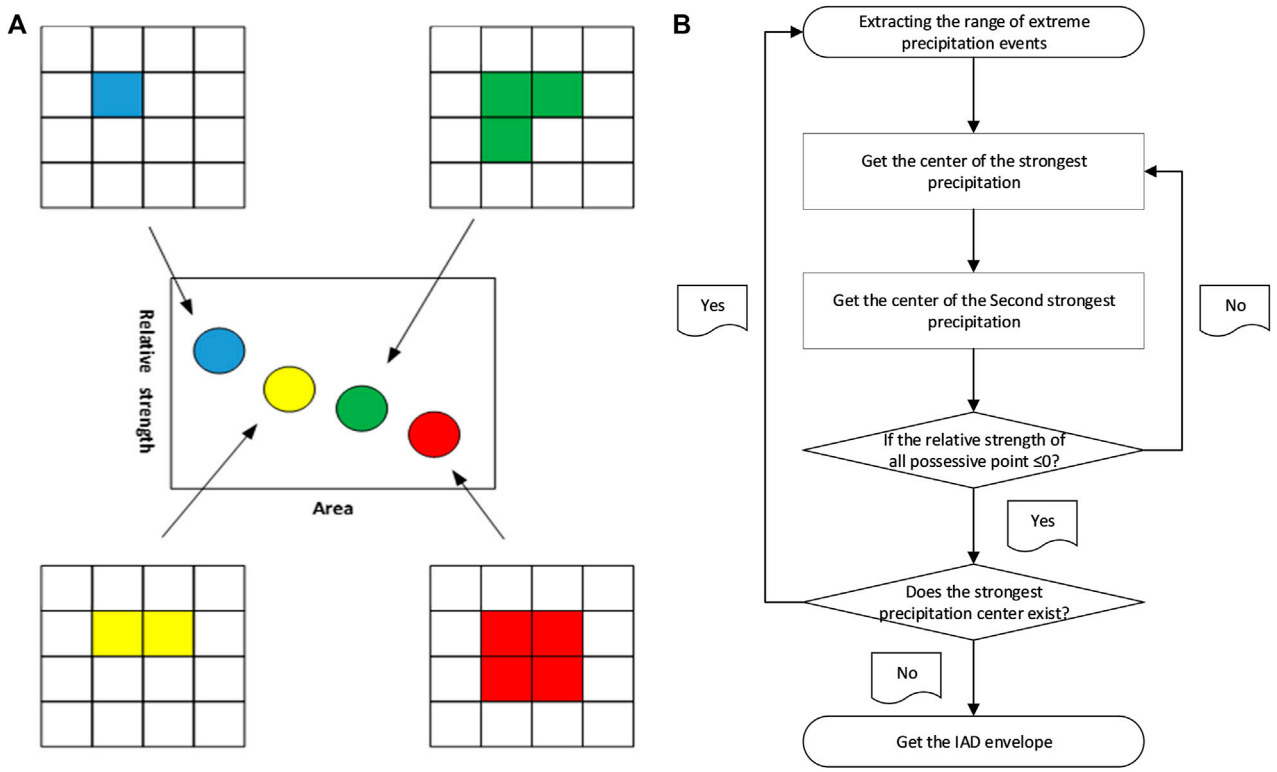


FIGURE 2 | Schematic diagram of identifying extreme precipitation events by the IAD method. **(A)** Distribution map of extreme precipitation events in different duration scales in the region. **(B)** Calculation steps of IAD identifying extreme precipitation events, which includes cyclic judgment calculation. The extreme precipitation events within the range are gradually identified according to the intensity until all identification is completed, and finally, the IAD envelope is obtained.

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}, \quad (1)$$

where n is the evaluation quantity of the same index, X_i is the *i*th value of the same index, and $\bar{X}$ is the average value of the index value.

The improved combination weighting method of game theory (ICWGT) analyzed the rationality and decision equilibrium of decision-making behavior when game theory interacts with each other by introducing game theory in the field of operations research. Its idea of combination weighting was to find a consistent or compromise weighting method among different weighting methods by minimizing the deviation between each index weight and the optimal linear combination index weight to achieve a balanced optimization method (Ren and Li, 2017) to screen the optimal combination weight. The combination weighting based on game theory could be expressed as follows:

$$w = \sum_{l=1}^{L} \alpha_l w_l^T, \quad (2)$$

where α_l is the linear combination coefficient, $\alpha_l > 0$, w is the combined weight vector, and w_l is the weight obtained by each weighting method. The weight vector w is combined with all $w_l$values; the objective was to minimize the deviation of L. By optimizing the L linear combination coefficients of the aforementioned formula, the optimal solution w* of w can be obtained. The resulting game model was as follows:

$$min\left\| \sum_{l=1}^{L} \alpha_l w_l^T - w_p \right\|_2, p = 1, 2..., L, \quad (3)$$

where *p* indicates the number of methods to calculate the weight of evaluation indicators, and the *p*th basic weight set is w_p.

In this study, the index weight of subjective weighting was obtained according to IFAHP, and the index weight of objective weighting was obtained according to CRITIC. On the basis of these, the combination coefficient was calculated through the improved game theory combination weighting, and the final index weight with combination weighting was normalized. Details of the process can be referred to Wu et al. (2021).

2.3.3 Accuracy Assessment

The global model data would inevitably appear in the simulation of regional precipitation, and there would be corresponding deviations in interpolating the grid data. To improve the simulation accuracy of the model data, a statistical downscaling correction method was used for model correction. Statistical downscaling of climate models was carried out through EDCDFm. It corrected the deviation of GCM-simulated climate elements through the difference of cumulative distribution

characteristics between measured data and GCM-simulated data to make the model achieve a more accurate simulation effect. It was assumed that the difference between the cumulative distribution probabilities of the two data in the observation stage would remain unchanged in the future.

$$\boldsymbol{F}(\boldsymbol{x}) = (\boldsymbol{1} - \boldsymbol{Q})\boldsymbol{H}(\boldsymbol{x}) + \boldsymbol{G}(\boldsymbol{X}), \tag{4}$$

where F(x) is the cumulative distribution function of precipitation in the observation period, and q is the proportion of precipitation months. H(x) is the step function. The month without precipitation is 0, and the month with precipitation is 1.

$$\boldsymbol{P}_{mcf} = \boldsymbol{F}_{OC}^{-1}(\boldsymbol{F}_{mc}(\boldsymbol{P}_{mc})), \tag{5}$$

$$\boldsymbol{P}_{mpj} = \boldsymbol{P}_{mp}\frac{\boldsymbol{F}_{oc}^{-1}\left(\boldsymbol{F}_{mp}\left(\boldsymbol{P}_{mp}\right)\right)}{\boldsymbol{F}_{mc}^{-1}\left(\boldsymbol{F}_{mp}\left(\boldsymbol{P}_{mp}\right)\right)}, \tag{6}$$

where P_{mcj} is the corrected value of the model data in the base period, P_{mpj} is the corrected value of the model data in the future period, $F_{\mathrm{oc}}{}^{-1}$ is the quantile function of the observed value in the base period, F_{mc} is the cumulative distribution function in the historical period of the model, P_{mc} is the precipitation data in the historical period of the model, P_{mp} is the precipitation data simulated in the future of the model, and $F_{\mathrm{mc}}{}^{-1}$ is the quantile function in the historical period of the model, and F_{mp} is the cumulative distribution function of data in the future period of the model.

RTGCV was selected to compare and analyze the model interpolation results with the observed values and combined with the root mean square error (RMSE) as the index to evaluate the interpolation effect. RMSE was the estimated value error after excluding the observed value error. The smaller the RMSE was, the better the interpolation effect was. Through verification, it was found that the fluctuation of RTGCV had apparent periodic law, with larger in summer and less in autumn and winter, and there were no significant interannual variation characteristics.

3 RESULTS

3.1 Data Accuracy Assessment in CPEC

For the validation of observed precipitation data sets used in this study, the statistical analysis showed that the annual average RMSE of interpolation grid point was 0.9 mm, which showed that the precipitation grid-point data had good accuracy and interpolation effect. The precipitation data obtained by GCMs were then compared with observed data sets. First, the simulation ability of the four models' data interpolated to the same accuracy was evaluated. Then, the multiyear average monthly precipitation was used as the evaluation index.

According to the existing research, the precipitation seasons in CPEC are from July to September. In particular, there are many extreme precipitation events in August, and the probability of extreme precipitation events will increase in the future (Bhatti et al., 2020). Therefore, the results obtained from the accurate evaluation of the model in August are more reasonable and representative. Consequently, the data of a grid point in August were taken as an example.

Figure 3A showed that the rainfall after correction matches well, which significantly reduced the correction error; **Figure 3B** showed the measured multiyear average monthly precipitation in the benchmark period 1984–2013 and the multiyear average monthly precipitation in the historical period of model simulation. It could be seen that except that MIROC6 obviously underestimates precipitation; most of the precipitation simulated by other models was slightly overestimated. Among them, the precipitation simulated by CNRM-CM6-1 was slightly underestimated by 10 mm in July, and the rest was overestimated somewhat, and its simulation situation was the closest. On the other hand, IPSL-CM6A-LR had the best simulation effect in March and May, and several models had significantly overestimated the simulation in November. In **Figure 3C**, the Taylor diagrams of four models were given, and the results of several models were in a good interval, among which CESM2 and CNRM-CM6-1 had smaller RMSD (Equivalent to RMSE divided by the standard deviation of the observed data).

Here, CESM2 and CNRM-CM6-1 with better precipitation simulation effect were selected for EDCDFm deviation correction, the simulation ability of the revised model was evaluated, and the Taylor diagram was used to compare the proximity between the two models and the observation data. The results of the model evaluation are shown in **Figure 3D**. Among them, the spatial correlation coefficients of the two corrected models were more outstanding than 0.9, the RMSD of CNRM-CM6-1 was smaller, and the ratio of their standard deviation was close to 1. Therefore, the model data simulation ability of CNRM-CM6-1 were more robust, and the three paths of the model data were selected for extreme precipitation event evaluation.

3.2 Projection of Extreme Precipitation in CPEC

For extreme precipitation events in different durations, the frequency difference of extreme precipitation events in the three scenarios in the future was not obvious under the condition of continuous 3d extreme precipitation.

Figure 4 shows the frequency of extreme precipitation events in the three scenarios in the future. The years of maximum frequency indicated under the three scenarios were different. Under SSP3-7.0, the frequency of extreme precipitation was the highest around 2037, and SSP1-2.6 was consistent with the year of the maximum frequency of extreme precipitation under SSP5-8.5. After reaching the maximum frequency, it showed a downward trend and then rose again after reaching the bottom in 2046. From the overall direction, under the three scenarios, the frequency of extreme precipitation fluctuated and increased and increased significantly in 2030. The interdecadal frequency variabilities under the three scenarios were 3.215 times/10A, 1.215 times/10A, and 5.063 times/10A, respectively. The interdecadal variability of extreme precipitation under the three scenarios was quite different. The

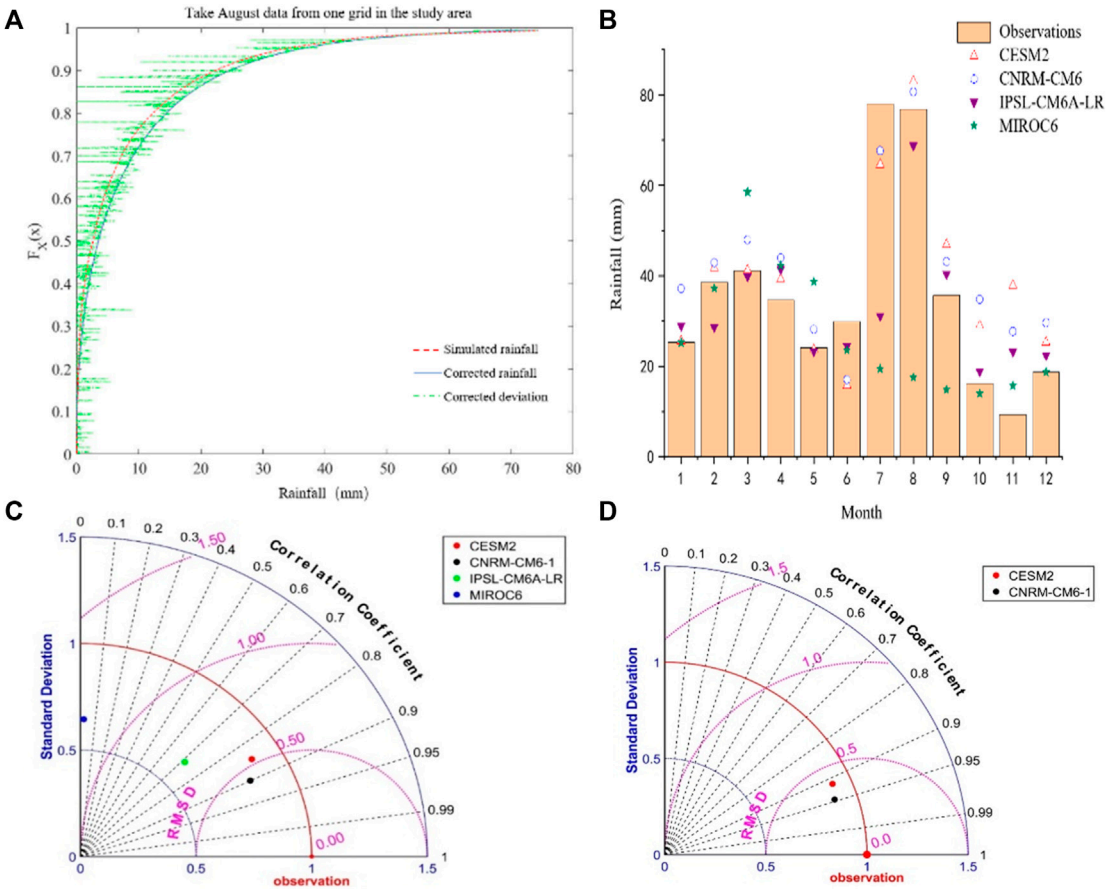


FIGURE 3 | (A) Model correction process (Taking August data of a grid point as an example). **(B)** Monthly precipitation in the base period and the monthly precipitation simulated by the model data. **(C)** Taylor diagram of four model data. **(D)** Taylor diagram of the model after deviation correction.

interdecadal variability under SSP5-8.5 was relatively large, in line with the climate change characteristics of high forcing and high radiation.

Comparing the occurrence of extreme precipitation events in the prediction stage with the average extreme precipitation frequency in the reference period, **Figure 5** was obtained. Compared with the base period, the extreme precipitation frequency under the three scenario models showed an upward trend, with significant extreme points under SSP1-2.6 and a large variation range under SSP3-7.0. On the other hand, under SSP5-8.5, the change of extreme precipitation frequency was relatively average, but it was always a large stage.

The impact area of extreme precipitation events was the grid area covered by an extreme event, and it was one of the important indicators to evaluate extreme precipitation events. In the prediction period, the average annual impact area under the three scenarios of extreme precipitation events lasting for 1 day generally showed an upward trend year by year (**Figure 6**). Among them, the impact area under SSP5-8.5 increased the fastest, and the minimum area exceeded 20000 km^2, the multiyear average impact area of a single event was 41000 km^2, and the maximum impact area reached 52.53 million km^2. As a result, the total impact area in the prediction period was 120.33 million km^2, and the total impact area in the benchmark period was 101.44 million km^2.

In **Figures 4–6**, we concluded that in the next 15 years, the average impact area of extreme precipitation events under the three scenarios would decrease, and the total impact area would increase to a certain extent. In addition, the number of extreme events would decrease slightly, indicating that the number of small-scale and high-intensity extreme precipitation events would increase from 2021 to 2035. In the next 30 years, the average impact area of extreme precipitation events would decrease, the total impact area would increase, and the number of extreme events would increase, indicating that the number of small-scale extreme precipitation events would increase from 2021 to 2050.

The observation of precipitation extremes was largely different among different SSPs. In identifying IAD extreme precipitation events, this study adopts the concepts of grid precipitation threshold and relative intensity. Most of the grid precipitation thresholds have increased in varying degrees under the following three scenarios, especially for periods of 5 days and 7 days. The multiyear monthly average precipitation was used as the standard to obtain the

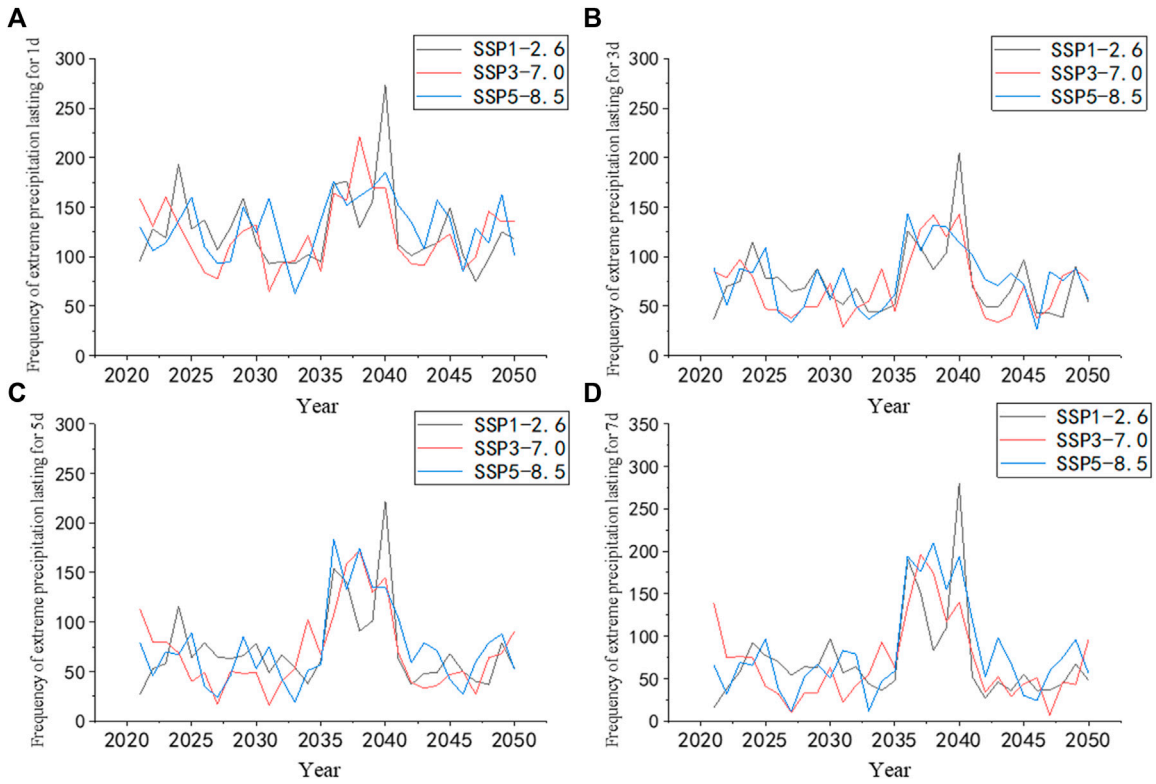


FIGURE 4 | Annual variation of the extreme precipitation frequency in three future scenarios under different durations. **(A)** Lasting for 1 day. **(B)** Lasting for 3 days. **(C)** Lasting for 5 days. **(D)** Lasting for 7 days.

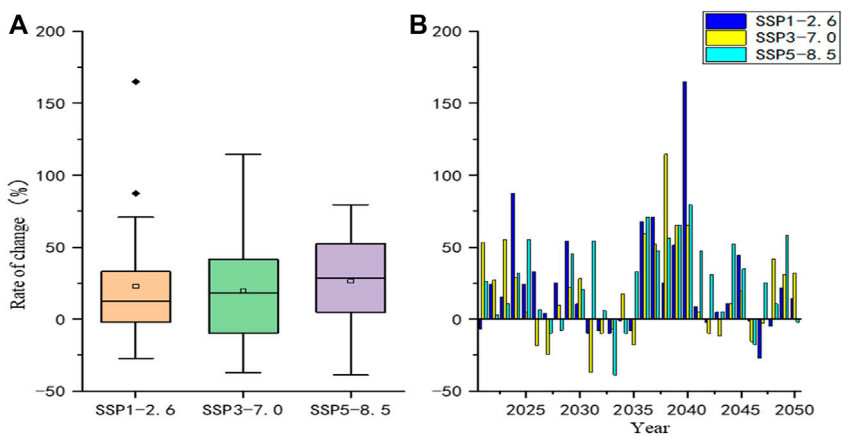


FIGURE 5 | Change percentage of extreme precipitation frequency under different scenarios in the future. **(A)** Average change percentage. **(B)** Change percentage of each year.

variation trend of future precipitation. **Figure 7** shows the multiyear average monthly precipitation during the observation period from 1984 to 2013 and the future three scenarios from 2021 to 2050. For the dry season, the precipitation under SSP1-2.6 was more than that under other paths, and the simulation of precipitation in SSP1-2.6 focused on balancing the precipitation in the dry season; In the rainy season (July/August), the simulation under SSP5-8.5 was more significant than the other two. This was a precipitation process simulated by high radiation and high forcing without climate policy intervention, which could better represent an extreme scenario. Therefore, the following inundation simulation and flood risk assessment were carried out using the precipitation data under this scenario.

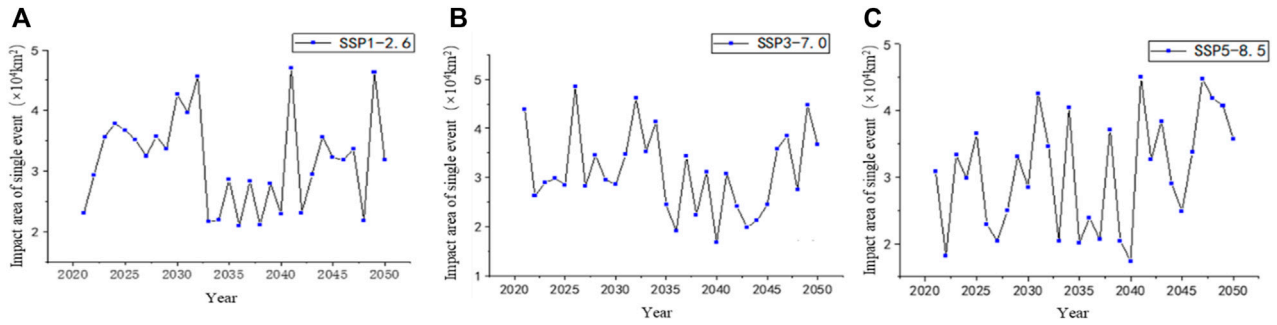


FIGURE 6 | Average annual impact area of extreme precipitation under different scenarios in the future. **(A)** Under SSP1-2.6. **(B)** Under SSP3-7.0. **(C)** Under SSP5-8.5.

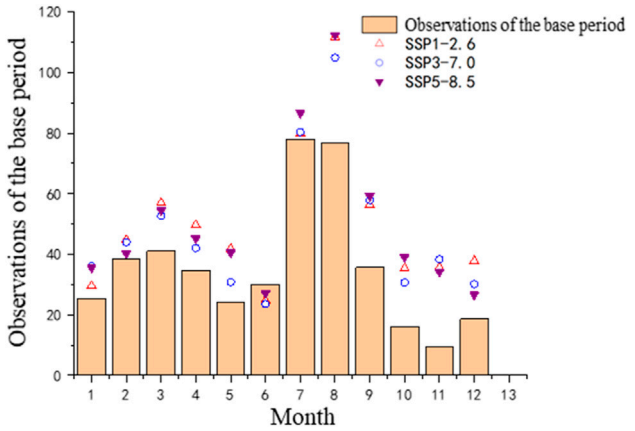


FIGURE 7 | Monthly average precipitation in the base period and monthly average precipitation in three combined paths from 2021 to 2050.

3.3 Projection of the Flood Risk in CPEC

In the context of global climate change, the intensity of extreme precipitation was increasing. Through the processing and research of model data, this study obtained that the extreme precipitation was the most obvious under SSP5-8.5. Therefore, the daily rainfall value in 2021–2050 under SSP5-8.5 and CNRM-CM6-1 was selected to calculate the area rainfall process with a 100-year return period and 20-year return period, and the corresponding DEM, land use, and other data were substituted into the flood area model for simulation, The flood inundation map (**Figure 8**) of the 100-year return period (high scenario) and 20-year (low scenario) under the future climate model scenario could be obtained to guide the flood prevention work under the future climate change scenario.

As shown in the figures, compared with the design flood inundation distribution in the historical period, the design flood inundation range had a partial increasing trend in the

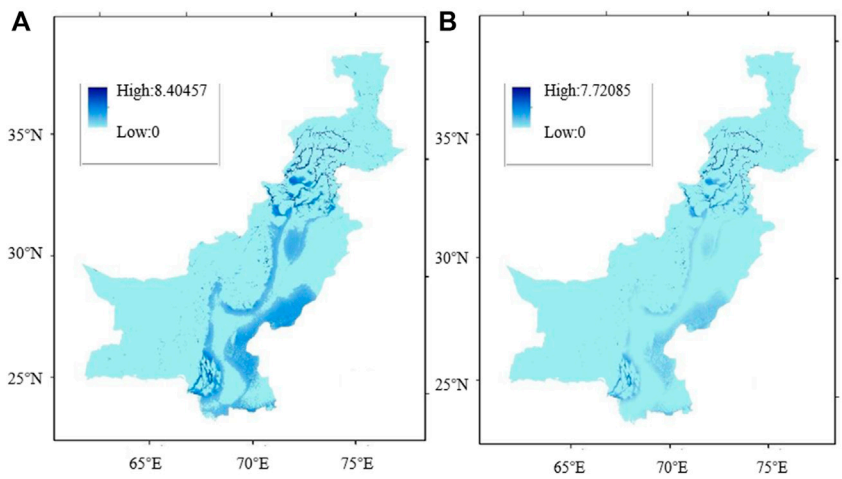


FIGURE 8 | Schematic diagram of the flood risk under the future scenario (SSP5-8.5). **(A)** 100-year flood process value. **(B)** 20-year flood process value.

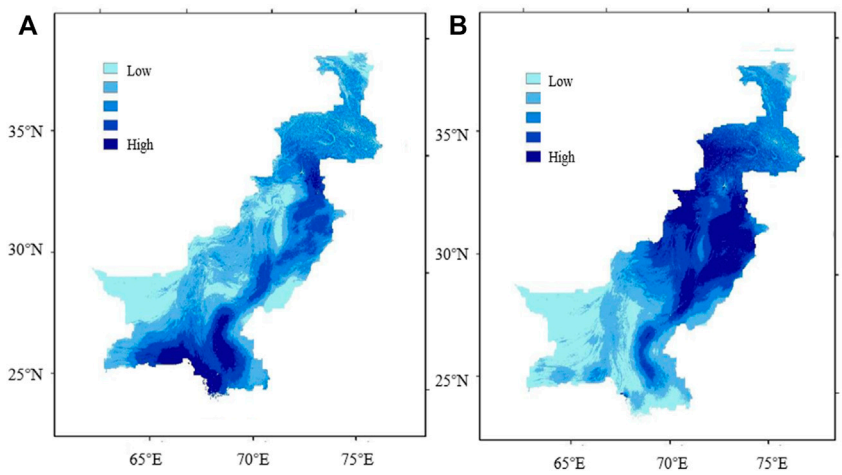


FIGURE 9 | Flood risk zoning in different return periods under future scenarios (SSP5-8.5). **(A)** 20-year flood risk zoning. **(B)** 100-year flood risk zoning.

future scenario and had increased in a small range. According to the statistics of inundation areas in different water depths, it was found that the increased range of inundation areas in different risk areas was different. Compared with the flood inundation in the historical period, the expansion area of the design 100-year flood under 1 m inundation depth was more prominent in the future scenario, and the expansion area of high water depth inundation was less pronounced.

As for frequency calculation, the daily rainfall data from 2021 to 2050 under SSP5-8.5 and CNRM-CM6-1 were selected for analysis. The grid-point rainfall was calculated as the design rainfall under the return period of 20 years and 100 years, respectively. The spatial distribution of design rainfall was obtained by interpolation of design rainfall to carry out the following risk assessment calculation. Finally, the 20-year flood risk distribution and 100-year flood risk distribution under the future scenario could be obtained (**Figure 9**). Compared with the historical observation period, under the future scenario, the area of medium- and high-risk areas in the 20-year flood risk zoning was partially expanded, especially located in the south of Sindh province and part of the Indus River into the sea, and the medium- and high-risk areas tend to expand in the middle of Punjab province. The expansion of the high-risk area with 100-year flood risk was the most obvious, which was mainly located in the plain of Punjab province. The area of the high-risk area was expanded by about 3.1%. The area of the low-risk area was in the south of Sindh province, and the south of Balochistan province had increased by 4.2%, and the area of the low-risk area and the medium-risk area had decreased correspondingly.

In general, the area of high-risk areas in the future scenario had an expanding trend, especially located in the east and south, affected by the summer monsoon and densely populated areas. Therefore, it was necessary to improve the flood prevention level further to avoid more significant losses in the future.

4 DISCUSSION

Due to the differences in simulation mechanism, topographic elements, and spatial resolution of different climate models, the simulation results were uncertain, and the simulation ability of other regions was different, especially in areas with complex topographic and atmospheric elements. The prediction of global precipitation data by the model data before CMIP5 was more in line with the measured situation than in a single region. In recent years, significant progress has been made in the regional simulation of GCMs. Huang et al. (2015) found that CMIP5 multimodel set data had a strong simulation ability for temporal and spatial temperature and precipitation changes. The model could also well simulate the seasonal fluctuations of precipitation. Chen et al. (2014) used 43 GCMs to predict the precipitation in China. They found that the CMIP5 model data could better simulate the regional distribution characteristics, which were higher in the southwest. Therefore, when using the global climate model to analyze the various features of meteorological elements under regional future climate change, it was necessary to select appropriate GCMs and evaluate the simulation ability of the climate model. Significantly, there would still be uncertainty in the application of GCM in predicting extreme regional events even if there were downscaling methods and deviation correction methods and the RCM dynamical downscaling method was still dominant in terms of regional precipitation simulation (Guo and Wang, 2016). In future research, it is suggested to use RCM or the GCM model with high accuracy and project precipitation extremes/floods.

Unlike CMIP5 model data, CMIP6 combined the typical concentration path and shared economy path to form a new scenario path model (SSP-RCP) to obtain more reliable prediction results. Jiang et al. (2020) also attempted to use CMIP6 data to evolve temporal–spatial characteristics of temperature and precipitation. However, IAD was not applied to their study. The climate in different regions of CPEC was quite different. Therefore, it was difficult to study the change

characteristics of extreme events in time series on a large unified scale and take the characteristics of different regions into account. The study area could be divided according to climatic and topographic characteristics to further explore the changing trend of extreme events in different regions. Spatially, the number of people affected by drought would be greater than that in the reference period. The increase in temperature exacerbated the drought. Regional drought risk levels were different (Wen et al., 2019), which proved IAD had a good effect on spatial characteristics. Considering the three-dimensional aspects of extreme precipitation events, IAD clustering was used to identify extreme precipitation events in CPEC in this study. The changing trend of extreme precipitation events under the background of climate change was obtained. It could expand the simulation path of the existing research to carry out more representative research on extreme climate events.

In this study, the newly released model data in CMIP6 were downscaled, the simulation ability was evaluated, and the applicability of CNRM-CM6-1 model data in CPEC was obtained. Compared with CMIP5, the single CNRM-CM6-1 model in CMIP6 could achieve the fitting effect of the multimodel aggregation average in CMIP5. Furthermore, the path data of three SSPs under CNRM-CM6-1 in CMIP6 were used to predict and evaluate extreme events under different climate change backgrounds in the future. Meanwhile, IFAHP, CRITIC, and ICWGT used in this study for assessing flood risk could make results obtained quickly, which was helpful for the division of flood risk areas. In previous studies, Abbas et al. (2022) found that under the high-forcing scenario (SSP5-8.5), the trend of extreme precipitation events in CPEC increased significantly, and the tendency of extreme precipitation events in summer also increased significantly, which is consistent with the conclusions of this article. In the verification of temperature simulation in Thailand (Suchada et al., 2021), the simulation results and accuracy of CNRM-CM6-1 were affirmed, so the accuracy of the simulation results based on CNRM-CM6-1 could be considered to be guaranteed. However, the resolution of CNRM-CM6-1 may still impact the accuracy of research results. Therefore, in future research, if the model's resolution can be solved, it will be an important breakthrough in improving the accuracy of disaster prediction.

5 CONCLUSION

For the rainfall data under the new combination scenario in the newly released CMIP6 model data, EDCDFm was used to correct the downscaling deviation. Before the correction, the model data were partially overestimated for rainfall. We selected the model closest to the measured data for correction. The single corrected model had a good fitting ability for the measured data's seasonal fluctuation and spatial distribution. The appropriate level of a single model could reach the level of the multimodel set in CMIP5.

Variation characteristics of extreme precipitation events under the background of climate change. Under CNRM-CM6-1, the frequency of extreme precipitation events presents interdecadal fluctuations of 3.215 times/10A, 1.215 times/10A, and 5.063 times/10A under three combined path datasets (SSP1-2.6, SSP3-7.0, and SSP5-8.5). In the next 30 years, the average impact area of extreme precipitation events would decrease, the total impact area would increase, and the extreme precipitation events in a small range would increase. Under the future scenario, the increased rate of extreme precipitation in August was the fastest, which increased the probability of extreme events.

For the flood risk under different return periods in the future, compared with the observation period, the flood risk had a more obvious expansion trend in the next 30 years, which was mainly reflected in the expansion of the area of high-, medium-, and low-risk areas. The risk zoning results obtained by the two different flood risk assessment methods differed, but the overall risk trend was the same.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material; further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

Conceptualization, SD; data curation, RW; formal analysis, JX and WL; methodology, YT; resources, WZ; supervision, HS; validation, RW; writing—original draft, SD and RW; writing—review and editing, HS, SD, and DY. All authors have read and agreed to the published version of the manuscript.

FUNDING

This research was primarily funded by the Ministry of Science and Technology of China (2019FY00205) (HS). The authors acknowledge funding from the NSFC project (51879110, 52079055, and 52011530128). HS acknowledges funding from the NSFC-STINT project (No. 202100-3211), and JX also acknowledges funding from the Youth Innovation Promotion Association of the Chinese Academy of Sciences (2019430).

REFERENCES

Abbas, A., Ullah, S., and Ullah, W. (2022). Evaluation and Projection of Precipitation in Pakistan Using the Coupled Model Intercomparison Project Phase 6 Model Simulations. *Int. J. Climatol.*, 1–20. doi:10.1002/joc.7602

Ali, A. F., Xiao, C.-D., Zhang, X.-P., Adnan, M., Iqbal, M., and Khan, G. (2018). Projection of Future Streamflow of the Hunza River Basin, Karakoram Range (Pakistan) Using HBV Hydrological Model. *J. Mt. Sci.* 15, 2218–2235. doi:10.1007/s11629-018-4907-4

Andreadis, K. M., Clark, E. A., Wood, A. W., Hamlet, A. F., and Lettenmaier, D. P. (2005). Twentieth-Century Drought in the

Conterminous United States. *J. Hydrometeorol.* 6 (6), 985–1001. doi:10.1175/jhm450.1

Benito, G., Macklin, M. G., Cohen, K. M., and Herget, J. (2015). Past Hydrological Extreme Events in a Changing Climate. *Catena* 130, 1–2. doi:10.1016/j.catena.2014.12.001

Bhatti, A. S., Wang, G., Ullah, W., Ullah, S., Fiifi Tawia Hagan, D., Kwesi Nooni, I., et al. (2020). Trend in Extreme Precipitation Indices Based on Long Term *In Situ* Precipitation Records Over Pakistan. *Water* 12, 797. doi:10.3390/w12030797

Boucher, O., Servonnat, J., and Albright, A. L. (2020). Presentation and Evaluation of the IPSL-CM6A-LR Climate Model. *J. Adv. Model. Earth Syst.* 12, e2019MS002010. doi:10.1029/2019MS002010

Brendel, C. E., Dymond, R. L., and Aguilar, M. F. (2021). Modeling Storm Sewer Networks and Urban Flooding in Roanoke, Virginia, with SWMM and GSSHA. *J. Hydrologic Eng.* 26 (1), 05020044. doi:10.1061/(asce)he.1943-5584.0002021

Chang, J., Gong, L., and Zeng, F. (2022). Using Hydro-Climate Elasticity Estimator and Geographical Detector Method to Quantify the Individual and Interactive Impacts on NDVI in Oasis-Desert Ecotone. *Stoch Stoch. Environ. Res. Risk Assess.*, 1–18. doi:10.1007/s00477-022-02184-4

Chen, X. C., Xu, Y., and Xu, C. H. (2014). Evaluation of CMIP5 Global Climate Model on Precipitation Simulation Ability in China. *Prog. Clim. Change Res.* 10 (03), 217–225. doi:10.3969/j.issn.1673-1719.2014.03.011

Danabasoglu, G., Lamarque, J.-F., and Bacmeister, J. (2020). The Community Earth System Model Version 2 (CESM2). *J. Adv. Model. Earth Syst.* 12, e2019MS001916. doi:10.1029/2019ms001916

Diakoulaki, D., Mavrotas, G., and Papayannakis, L. (1995). Determining Objective Weights in Multiple Criteria Problems: The Critic Method. *Comput. Operations Res.* 22 (7), 763–770. doi:10.1016/0305-0548(94)00059-h

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., et al. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) Experimental Design and Organization. *Geosci. Model Dev.* 9 (5), 1937–1958. doi:10.5194/gmd-9-1937-2016

Federal Flood Commission, Ministry of Water and Power (2015). *Annual Flood Report 2015*. Islamabad (Pakistan): Government of Pakistan Ministry of Water and Power.

Goswami, B. N., Venugopal, V., Sengupta, D., Madhusoodanan, M. S., and Xavier, P. K. (2006). Increasing Trend of Extreme Rain Events Over India in a Warming Environment. *Science* 314 (5804), 1442–1445. doi:10.1126/science.1132027

Guo, D., and Wang, H. (2016). Comparison of a Very-Fine-Resolution GCM with RCM Dynamical Downscaling in Simulating Climate in China. *Adv. Atmos. Sci.* 33, 559–570. doi:10.1007/s00376-015-5147-y

Haq, M., Akhtar, M., Muhammad, S., Paras, S., and Rahmatullah, J. (2012). Techniques of Remote Sensing and GIS for Flood Monitoring and Damage Assessment: A Case Study of Sindh Province, Pakistan. *Egypt. J. Remote Sens. Space Sci.* 15 (2), 135–141. doi:10.1016/j.ejrs.2012.07.002

Huang, J. L., Su, B. D., and Zhu, X. Y. (2015). Simulation and Prediction of Climate Change in the Indus River Basin in South Asia by CMIP5 Multi-Model Set. *Glacial Permafr.* 37 (02), 297–307. doi:10.7522/j.issn.1000-0240.2015.0033

Huang, J., Zhang, X., Zhang, Q., Lin, Y., Hao, M., Luo, Y., et al. (2017). Recently Amplified Arctic Warming Has Contributed to a Continual Global Warming Trend. *Nat. Clim. Change* 7 (12), 875–879. doi:10.1038/s41558-017-0009-5

Jiang, T., Lv, Y. R., and Huang, J. L. (2020). Overview of CMIP6 Model New Scenario (SSP-RCP) and its Application in Huaihe River Basin. *Adv. Meteorol. Sci. Technol.* 10 (05), 102–109. doi:10.3969/j.issn.2095-1973.2020.05.016

Jing, C., Jiang, T., and Wang, Y. J. (2016). Study on Regional Extreme Precipitation Events and Population Economic Exposure in China. *J. Meteorol.* 74 (04), 572–582. doi:10.11676/qxxb2016.037

Kataoka, T., Tatebe, H., and Koyama, H. (2020). Seasonal to Decadal Predictions with MIROC6: Description and Basic Evaluation. *J. Adv. Model. Earth Syst.* 12, e2019MS002035. doi:10.1029/2019ms002035

Khan, S. I., Hong, Y., Gourley, J. J., Khattak, M. U. K., Yong, B., and Vergara, H. J. (2014). Evaluation of Three High-Resolution Satellite Precipitation Estimates: Potential for Monsoon Monitoring over Pakistan. *Adv. Space Res.* 54, 670–684. doi:10.1016/j.asr.2014.04.017

Li, H., Chen, H., Wang, H., and Yu, E. (2018). Future Precipitation Changes over China under 1.5 °C and 2.0 °C Global Warming Targets by Using CORDEX Regional Climate Models. *Sci. Total Environ.* 640-641, 543–554. doi:10.1016/j.scitotenv.2018.05.324

Lu, M., Sun, H., Yan, D., Xue, J., Yi, S., Gui, D., et al. (2021). Projections of Thermal Growing Season Indices over China under Global Warming of 1.5 °C and 2.0 °C. *Sci. Total Environ.* 781, 146774. doi:10.1016/j.scitotenv.2021.146774

Meehl, G. A., Karl, T., Easterling, D. R., Changnon, S., Pielke, R., Changnon, D., et al. (2000). An Introduction to Trends in Extreme Weather and Climate Events: Observations, Socioeconomic Impacts, Terrestrial Ecological Impacts, and Model Projections. *Bull. Amer. Meteor. Soc.* 81 (3), 413–416. doi:10.1175/1520-0477(2000)081<0413:aittie>2.3.co;2

Memon, A. A., Muhammad, S., Rahman, S., and Haq, M. (2015). Flood Monitoring and Damage Assessment Using Water Indices: A Case Study of Pakistan Flood-2012. *Egypt. J. Remote Sens. Space Sci.* 18 (1), 99–106. doi:10.1016/j.ejrs.2015.03.003

O'Neill, B. C., Tebaldi, C., van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., et al. (2016). The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6. *Geosci. Model Dev.* 9 (9), 3461–3482. doi:10.5194/gmd-9-3461-2016

Pachauri, R. K., and Reisinger, A. E. (2008). *Climate Change 2007. Synthesis Report. Contribution of Working Groups I, II and III to the Fourth Assessment Report*. Geneva (Switzerland), Switzerland: IPCC.

Pierce, D. W., Barnett, T. P., Santer, B. D., and Gleckler, P. J. (2009). Selecting Global Climate Models for Regional Climate Change Studies. *Proc. Natl. Acad. Sci. U.S.A.* 106 (21), 8441–8446. doi:10.1073/pnas.0900094106

Ren, L. C., and Li, Z. F. (2017). A New Model Based on the Games Theory and Fuzzy Mathematics in Bridge Engineering Risk Assessment. *Highw. Eng.* 42 (1), 163–169. doi:10.3969/j.issn.1674-0610.2017.01.037

Riahi, K., van Vuuren, D. P., Kriegler, E., Edmonds, J., O'Neill, B. C., Fujimori, S., et al. (2017). The Shared Socioeconomic Pathways and Their Energy, Land Use, and Greenhouse Gas Emissions Implications: An Overview. *Glob. Environ. Change* 42, 153–168. doi:10.1016/j.gloenvcha.2016.05.009

Sadiq, R., and Tesfamariam, S. (2009). Environmental Decision-Making under Uncertainty Using Intuitionistic Fuzzy Analytic Hierarchy Process (IF-AHP). *Stoch. Environ. Res. Risk Assess.* 23 (1), 75–91. doi:10.1007/s00477-007-0197-z

Safi, U., You, Q. L., and Waheed, U. (2018). Observed Changes in Precipitation in China-Pakistan Economic Corridor during 1980–2016. *Atmos. Res.* 210, 1–14. doi:10.1016/j.atmosres.2018.04.007

Simpkins, G. (2017). Progress in Climate Modelling. *Nat. Clim. Change* 7 (10), 684–685. doi:10.1038/nclimate3398

Su, B. D., Jiang, T., and Dong, W. J. (2008). Statistical Fitting of Distribution Characteristics of Extreme Heavy Precipitation in the Yangtze River Basin. *Meteorol. Sci.* 28 (06), 625–629. doi:10.3969/j.issn.1009-0827.2008.06.006

Su, B. D., Jiang, T., and Ren, G. Y. (2006). Temporal and Spatial Variation Trend of Extreme Heavy Precipitation in the Yangtze River Basin from 1960 to 2004. *Prog. Clim. Change Res.* 2 (01), 9–14. doi:10.3969/j.issn.1673-1719.2006.01.002

Suchada, K., Pham, T. B. T., and Shabbir, H. G. (2021). Evaluation of CMIP6 GCMs for Simulations of Temperature Over Thailand and Nearby Areas in the Early 21st Century. *Heliyon* 7 (11), e08263. doi:10.1016/j.heliyon.2021.e08263

Sun, H., Bai, Y., Lu, M., Wang, J., Tuo, Y., Yan, D., et al. (2021). Drivers of the Water Use Efficiency Changes in China during 1982-2015. *Sci. Total Environ.* 799, 149145. doi:10.1016/j.scitotenv.2021.149145

Sun, H., Chen, L., Yang, Y., Lu, M., Qin, H., Zhao, B., et al. (2022). Assessing Variations in Water Use Efficiency and Linkages with Land-Use Changes Using Three Different Data Sources: A Case Study of the Yellow River, China. *Remote Sens.* 14 (5), 1065. doi:10.3390/rs14051065

Sun, W., Mu, X., Song, X., Wu, D., Cheng, A., and Qiu, B. (2016). Changes in Extreme Temperature and Precipitation Events in the Loess Plateau (China) during 1960-2013 under Global Warming. *Atmos. Res.* 168, 33–48. doi:10.1016/j.atmosres.2015.09.001

Voldoire, A., Saint-Martin, D., Sénési, S., Decharme, B., Alias, A., Chevallier, M., et al. (2019). Evaluation of CMIP6 DECK Experiments with CNRM-CM6-1. *J. Adv. Model. Earth Syst.* 11, 2177–2213. doi:10.1029/2019ms001683

Wang, D. G., Jiang, P., and Wang, G. L. (2015). Urban Extent Enhances Extreme Precipitation over the Pearl River Delta, China. *Atmos. Sci. Lett.* 16 (3), 310–317. doi:10.1002/asl2.559

Wang, G. L., Wang, D. G., and Kevin, E. T. (2017). The Peak Structure and Future Changes of the Relationships between Extreme Precipitation and Temperature. *Nat. Clim. Change* 7 (4), 268–274. doi:10.1038/nclimate3239

Wang, Y., Wan, Q., Meng, W., Liao, F., Tan, H., and Zhang, R. (2011). Long-term Impacts of Aerosols on Precipitation and Lightning over the Pearl River Delta Megacity Area in China. *Atmos. Chem. Phys.* 11, 12421–12436. doi:10.5194/acp-11-12421-2011

Wen, S. S., Wang, A. Q., and Tao, H. (2019). Population Exposed to Drought under the 1.5 °C and 2.0 °C Warming in the Indus River Basin. *Atmos. Res.* 218, 296–305. doi:10.1016/j.atmosres.2018.12.003

Wu, R. Y., Sun, H. W., and Yan, D. (2021). Evaluation of Flood Disaster Risk in China-Pakistan Economic Corridor by Combination Weighting Based on Improved Game Theory and Grid Data. *Trans. Chin. Soc. Agric. Eng. (Transactions CSAE)* 37 (14), 145–154. doi:10.11975/j.issn.1002-6819.2021.14.016

Xue, J. G., Mei, L. W., and Giorgi, F. (2013). Climate Change over China in the 21st Century as Simulated by BCC_ CSM1. 1-RegCM4. 0. *Atmos. Ocean. Sci. Lett.* 6 (05), 381–386. doi:10.1080/16742834.2013.11447112

Yang, L. C., Deng, S., and Xu, J. H. (2010). Flood Risk Assessment Algorithm Based on BP Network. *Comput. Technol. Dev.* 20 (04), 232–234. doi:10.3969/j.issn.1673-629X.2010.04.059

Yang, Y., Sun, H., Xue, J., Liu, Y., Liu, L., Yan, D., et al. (2021). Estimating Evapotranspiration by Coupling Bayesian Model Averaging Methods with Machine Learning Algorithms. *Environ. Monit. Assess.* 193, 156. doi:10.1007/s10661-021-08934-1

Zhang, L. P., Du, H., and Xia, J. (2011). Research Progress of Extreme Hydrological Events under Climate Change. *Prog. Geogr. Sci.* 30 (11), 1370–1379. doi:10.11820/dlkxjz.2011.11.006

Zhang, Q., Marco, G., and Chen, J. Q. (2008). Climate Changes and Flood/drought Risk in the Yangtze Delta, China, during the Past Millennium. *Quat. Int.* 176-177, 62–69. doi:10.1016/j.quaint.2006.11.004

Zhao, B. Q., Sun, H. W., and Yan, D. (2021). Quantifying Changes and Drivers of Runoff in the Kaidu River Basin Associated with Plausible Climate Scenarios. *J. Hydrol. Reg. Stud.* 38, 100968. doi:10.1016/j.ejrh.2021.100968

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Environmental Science

ORIGINAL RESEARCH
published: 06 July 2022
doi: 10.3389/fenvs.2022.887040

Water Conservation Ecological Service Function and Its Value Response Mechanism in a Nested Water Conservancy Project Area

Chunfen Zeng[1,2,3]*, *Wanyu Qi*[2], *Yuqing Mao*[2], *Rui Liu*[2,3,4]*, *Boya Yu*[2] and *Xinning Dong*[5]

[1]Key Laboratory of Surface Processes and Environment Remote Sensing in the Three Gorges Reservoir Area, Chongqing Normal University, Chongqing, China, [2]School of Geography and Tourism, Chongqing Normal University, Chongqing, China, [3]The Key Laboratory of GIS Application Research, School of Geography and Tourism, Chongqing Normal University, Chongqing, China, [4]National Earth System Science Data Center, Beijing, China, [5]Chongqing Climate Center, Chongqing, China

OPEN ACCESS

Edited by:
Zhenzhong Zeng,
Southern University of Science and Technology, China

Reviewed by:
Chuanfu Zang,
South China Normal University, China
Guoqing Wang,
Nanjing Hydraulic Research Institute, China

***Correspondence:**
Chunfen Zeng
cfzeng18@163.com
Rui Liu
liur@cqnu.edu.cn

Specialty section:
This article was submitted to Environmental Informatics and Remote Sensing, a section of the journal Frontiers in Environmental Science

Received: 01 March 2022
Accepted: 09 May 2022
Published: 06 July 2022

Citation:
Zeng C, Qi W, Mao Y, Liu R, Yu B and Dong X (2022) Water Conservation Ecological Service Function and Its Value Response Mechanism in a Nested Water Conservancy Project Area. Front. Environ. Sci. 10:887040. doi: 10.3389/fenvs.2022.887040

The Three Gorges Project is the largest water conservancy project in the world. To cope with the ecological problems of the subsidence zone in the Three Gorges Reservoir area, the Kaizhou water-level-regulating dam was built, forming a model of nested water conservancy projects. The Pengxi River Basin is affected not only by this project but also by human activities in the reservoir area and changes in the substratum and hydro-meteorology, which exert influences on the water connotation function and its ecological value through complex mechanisms. In addition, the response mechanism of the changes in the environment is unclear. Therefore, based on the Integrated Valuation of Ecosystem Services and Tradeoffs (InVEST) model and a spatial interpolation method, in this study, two time nodes (2005 and 2018) before and after the operation of the nested water conservancy projects began were selected, and seven simulation scenarios with different water levels, precipitations, and temperatures were created to explore the evolution of the water conservation service function in the nested water conservancy project operation area under the complex changes in the environment. The results reveal that the operation of the water conservancy projects has had some influence on the water content, but the response of the water content function to the precipitation conditions has been more significant. In colder and rainier years, the water content was higher. In contrast, the lowest value occurred in a year with high temperatures and low rainfall. The highest and lowest values were quantitatively different. Therefore, the influences of the complex environmental factors on the regional water connotation service function deserve more attention. The results of this study provide a scientific basis for research on the ecological service function and the value of water conservation in the Three Gorges Reservoir subsidence zone and the nested operation area of the related water conservancy projects, as well as a data reference for the optimal allocation of regional water resources.

Keywords: InVESTt model1, water yield2, nested area of water conservancy project3, hydrology weather condition4, Lucc5

1 INTRODUCTION

Under the context of climate change, ecological degradation has become a global priority. As a result, human well-being and livelihoods depend heavily on freshwater ecosystems and their service functions (Capon et al., 2018). Water conservation service is an important link between natural ecosystems and human society's demand for ecosystem services, and it plays an important role in the water cycle and water balance of river basins, as well as human survival and development (Ren and Mao, 2021).

The integrated valuation of ecosystem services and tradeoffs (InVEST) model has been widely used in water conservation services research. This model has been used in many regions, such as the Chindwin River Basin in Myanmar, the Portuguese Continental Basin, the Sokoto–Rima Basin in West Africa, Chile in South America, and the middle and lower reaches of the Yangtze River in China (Almeida and Cabral, 2021; Benra et al., 2021; Raji et al., 2021; Shrestha et al., 2021; Chen et al., 2022). This model can be used not only to evaluate the ecological service functions of different regions but also to predict the changes in ecological services by simulating different scenarios in order to identify the best scenario for sustainable regional development.

In recent years, due to the impacts of global climate change and social and economic development, Chinese and foreign scholars have focused on analyzing the driving factors of regional water conservation services with the help of this model. Dai and Wang (2020) analyzed the spatial heterogeneity and conducted the attribution analysis of water producing services in the Hengduan mountainous area and found that climate factors were the main factors controlling the spatial heterogeneity of the water producing service. In addition, evapotranspiration and precipitation can be interpreted differently for different topography. Wang et al. (2020) used the InVEST model to simulate the water conservation scenarios for the Fujian Triangle urban agglomeration in 2015 and 2030 and found that the impact of land-use changes on water conservation is mainly manifested in four aspects: changes in area, changes in direction, action intensity, and area compensation. Aneseyee et al. (2022) used the InVEST model to analyze the spatial and temporal variabilities of the water yield under land-use change (LUC) and climate change in the Winike watershed, Ethiopia, and found that land conversion affected the water yield but precipitation had a greater impact. Wang et al. (2022) quantitatively assessed the spatial and temporal variabilities of the ecosystem services in the Qilian Mountains using the InVEST model, as well as the differences in the eco-hydrological services under different land-use types. The Daqinghe River Basin was selected as an example to study the changes in the ecosystem services and its influencing factors. The results revealed that the natural background conditions and ecological protection policies were the main factors driving the changes in the ecosystem regulation services. The supply of agricultural products and the change in the area of construction land were the main factors driving the spatiotemporal changes in the agricultural product supply service and human settlement environment, respectively (Pan et al., 2021). Li J. et al. (2021) studied the impacts of precipitation and land-use changes on the water supply and service function in the Beisanhe River Basin. Their results revealed that the impact of the climate on the water yield was far greater than that of the land-use changes, and the surge in precipitation played a leading role. Zhao et al. (2019) analyzed the driving factors affecting the water yield in the upper reaches of the Shiyang River and concluded that climate was the leading factor affecting the changes in the water yield. Most of these studies focused on the response of water harvesting to land-use changes and climate change. With the increasing disturbance of human activities, especially the construction and operation of large-scale water conservancy projects, the water resources and ecosystems in the basin are greatly affected. However, few studies have been conducted on the mechanism of water-level fluctuations under the influence of water conservancy projects and their effect on water conservation services.

The Three Gorges Project formed a fluctuation zone, and its ecosystem is fragile. Related studies have mainly focused on the health and change mechanism of the ecosystem in the water-level fluctuation zone of the Three Gorges Reservoir area, including the soil, vegetation, ecological problems, and restoration (He and Bao, 2019; Lu, 2021; Lu et al., 2021; Shen et al., 2021). In addition, the characteristics of the water-level fluctuation zone, such as its long and narrow shape, cloudy weather, and large variations in the water level, increase the difficulty of data acquisition, and thus, previous research studies on the water conservation services are mainly based on static evaluations (Zhang et al., 2007). Few studies have been conducted on the dynamic changes caused by the operation of nested water conservancy projects. The operation of large-scale water conservancy projects has significantly affected the regional water and soil resources and the ecological services (Cheng et al., 2015; Cheng et al., 2018; Hu, 2019; Li and Hao, 2019; Li et al., 2019; Gong et al., 2020; Li et al., 2021a). Lu et al. (2021) analyzed the comprehensive effect and driving mechanism of the ecological water transfer project in the Heihe River Basin since 2000 and found that the hydrological, ecological, and socio-economic systems in the LXS region have formed feedback loops, but the comprehensive correlation among the water resources, ecosystem, and socio-economic system should still be considered in order to coordinate the relationship between them.

The Pengxi River Basin was chosen as the research area of this study. A large water-level fluctuation zone has formed in this basin under the influences of the operation of the Three Gorges Project and the special terrain conditions in this region. It accounts for about 45% of the water-level fluctuation zone in the Three Gorges Reservoir area, which makes this watershed ecologically fragile. In order to reduce the deterioration of the ecological environment in the water-level fluctuation zone and its surrounding areas, a water-level-regulating dam was built in Kaizhou. The operation of nested water conservancy projects makes the regional ecological driving mechanism more complex. Under the multiple effects of land-use changes, the operation of nested water conservancy projects, the human activities in the reservoir area, and changes in the meteorological conditions, the

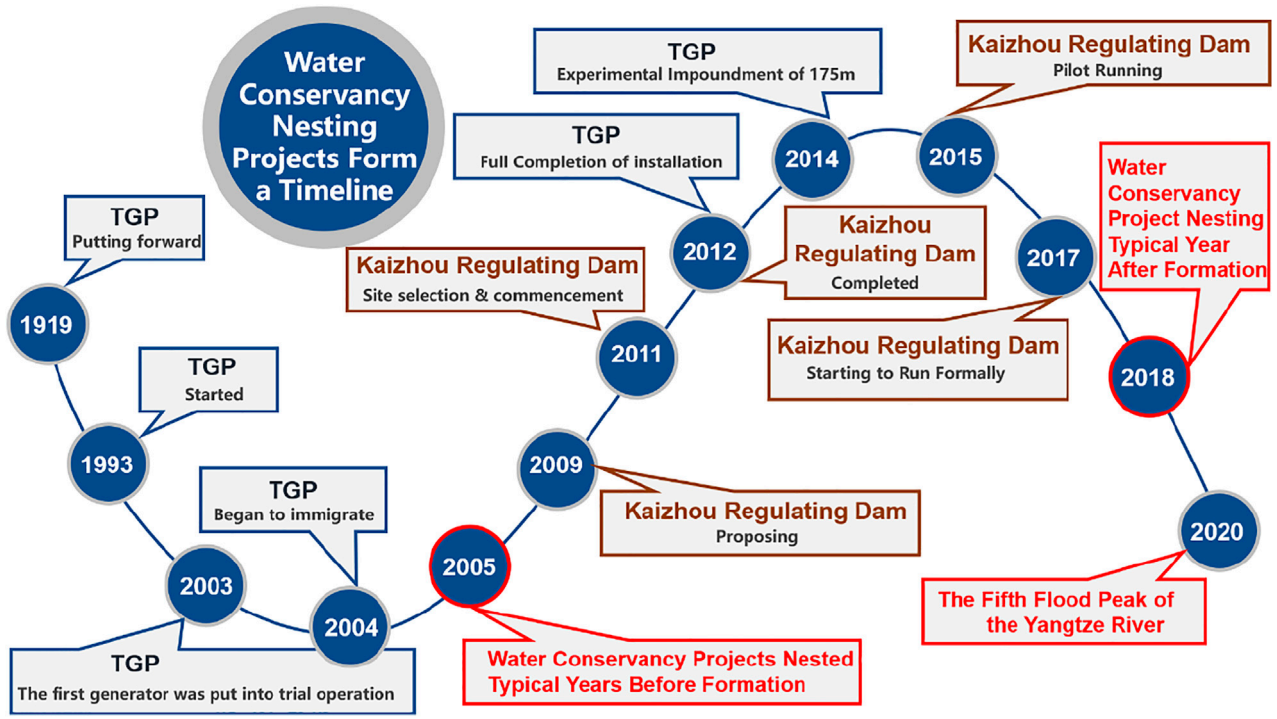


FIGURE 1 | Water conservancy projects are nested to form a timeline.

change mechanisms of the ecological service functions and the value of the water conservation need to be investigated further. Therefore, in this study, the Pengxi River Basin was taken as the study area, and the InVEST model was used to evaluate the water conservation service functions under the effects of multiple impacts via statistical analysis and geographic information system (GIS) technology and to determine the change mechanism of the water conservation ecological service functions under the influence of human activities, land-use changes, changes in the meteorological conditions, and water-level control of different water conservancy projects (the water-level control of the Three Gorges Project: low water level 145 m, high water level 175 m, the same below). The results of this study provide a scientific basis for ecological protection and restoration and the optimal allocation of regional water and soil resources in the Pengxi River Basin and the affected areas of related large-scale water conservancy projects and a theoretical reference for the ecological effects of related engineering measures.

2 MATERIALS AND METHODOLOGY

2.1 Study Area

The operation of the Three Gorges Project has formed a water fluctuation zone at 145–175 m in the reservoir. The special topography of the Pengxi River Basin in the Chongqing section of the Three Gorges Reservoir area has resulted in the formation of a large water-level fluctuation zone, accounting for about 45% of the Three Gorges Reservoir area. In order to improve the water-level fluctuation zone and the vulnerability of the surrounding ecological environment, the Kaizhou water-level regulation dam was constructed to form a nested area of large-scale water conservancy projects (**Figure 1**).

The Pengxi River Basin (**Figure 2**) is located in the eastern part of the Sichuan Basin (107°56'–108°54'E, 30°49'–31°42'N), and the Pengxi River mainly flows through the Kaizhou District and Yunyang County in Chongqing, with a total basin area of 5,173 m^2. The main landforms in the basin include mountains, hills, and plains. The topography of the region is undulating and is generally high in the northeast and low in the southwest. Most of the region has a subtropical humid monsoon climate, except for the northern mountainous area, which has an altitude of more than 1,000 m and a warm temperate monsoon climate.

The average annual temperature is 18.47–10.81°C, and the temperature difference between the northern and southern parts of the study area is large due to the influence of the vertical temperature gradient. The rainfall is concentrated in summer and autumn, and the rainy season is long. The annual average rainfall is 1,100–1,500 mm, and the annual average runoff is 3.58×10^9 m^3. The forest coverage rate of the basin is 31.4%, and the vertical zoning of the vegetation is obvious. Purple soil, yellow soil, and paddy soil are the main soil types.

2.2 Research Data

The sources of the land cover, meteorological, and socio-economic data used in this study are presented in **Table 1**. In this study, two time nodes were selected: one before the operation

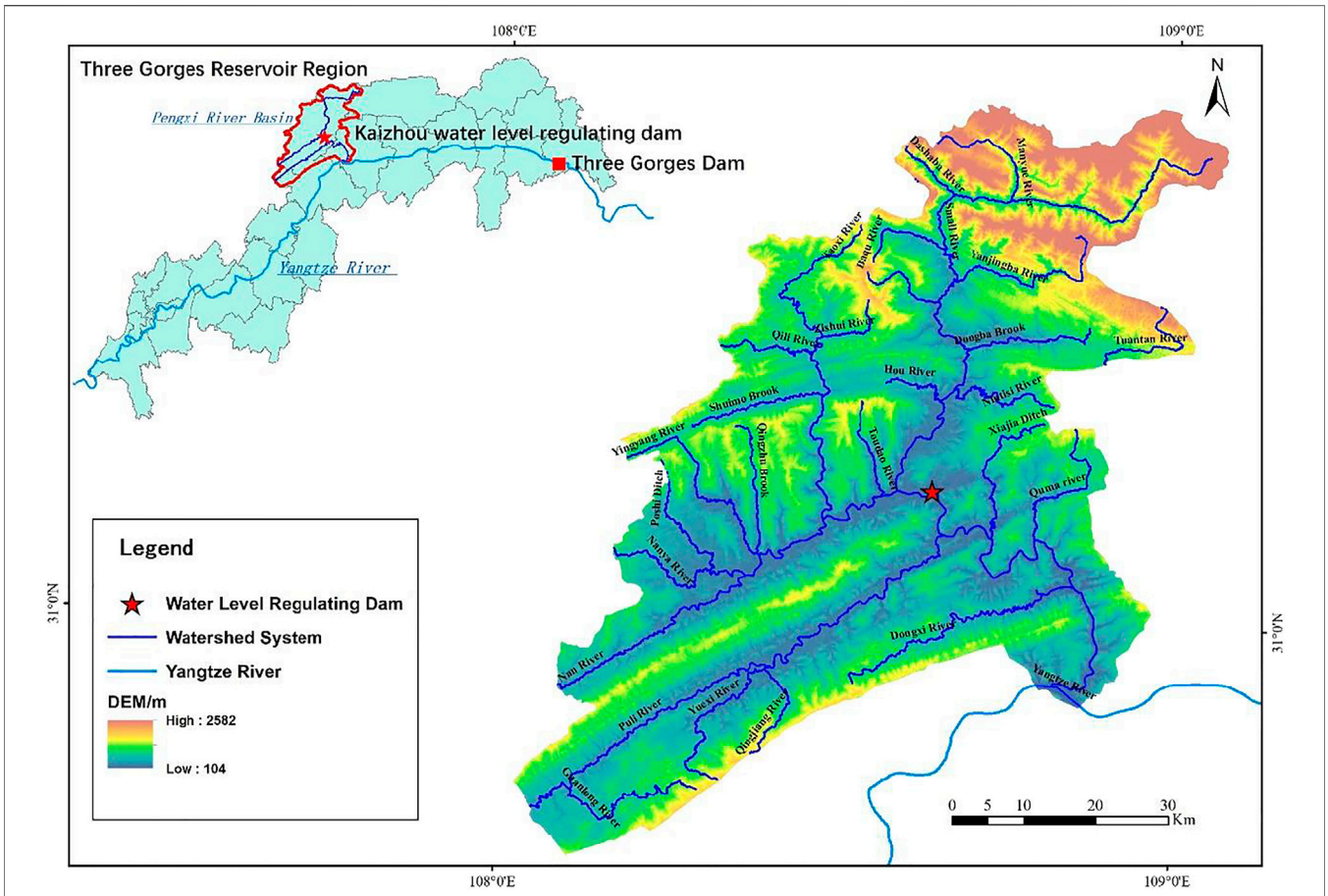


FIGURE 2 | Schematic diagram showing the location, elevation, water system, and dam in the study area.

TABLE 1 | Data and sources.

Data	Data Source
Land use	Southwest Mountain Center of the National Earth System Science Data Center
NDVI	IKONOS satellite and WorldView-3 satellite
DEM	ALOS-DSM satellite
Monthly and yearly precipitation	Southwest Mountain Center of the National Earth System Science Data Center
Daily temperature	National Meteorological Science Data Center
Radiation from the top of the solar atmosphere	Data publishing system for global change science research
Based on HWSD China soil dataset	National Earth System Science Data Platform—Cold Area and Early Area Data Center
Evapotranspiration coefficient	InVEST model database
Root depth of vegetation	InVEST model database
Velocity coefficient	InVEST model database
Socio-economic data	Statistical Yearbook and Bulletin of Chongqing Municipality and Kaizhou District of Chongqing Municipality

of the nested hydraulic project began (2005) and one after the operation of the nested hydraulic project began (2018). According to the land-use classification standard GB/T21010-2017, the spatial distributions of the ecosystems in 2005 and 2018 in the study area were determined. The hydrological analysis tool in ArcGIS was used to determine the boundary of the watershed in the study area based on a digital elevation model (DEM). Based on hydrological and meteorological data from three meteorological stations in the Kaizhou District and its surrounding areas for typical years, such as a wet year, a normal water year, a dry year, a high-temperature year, and a low-temperature year, an annual rainfall distribution map and temperature distribution map were generated via inverse distance weighted interpolation. Hydrometeorological simulation scenarios were defined based on the temperature, precipitation, and solar radiation data in order to calculate the

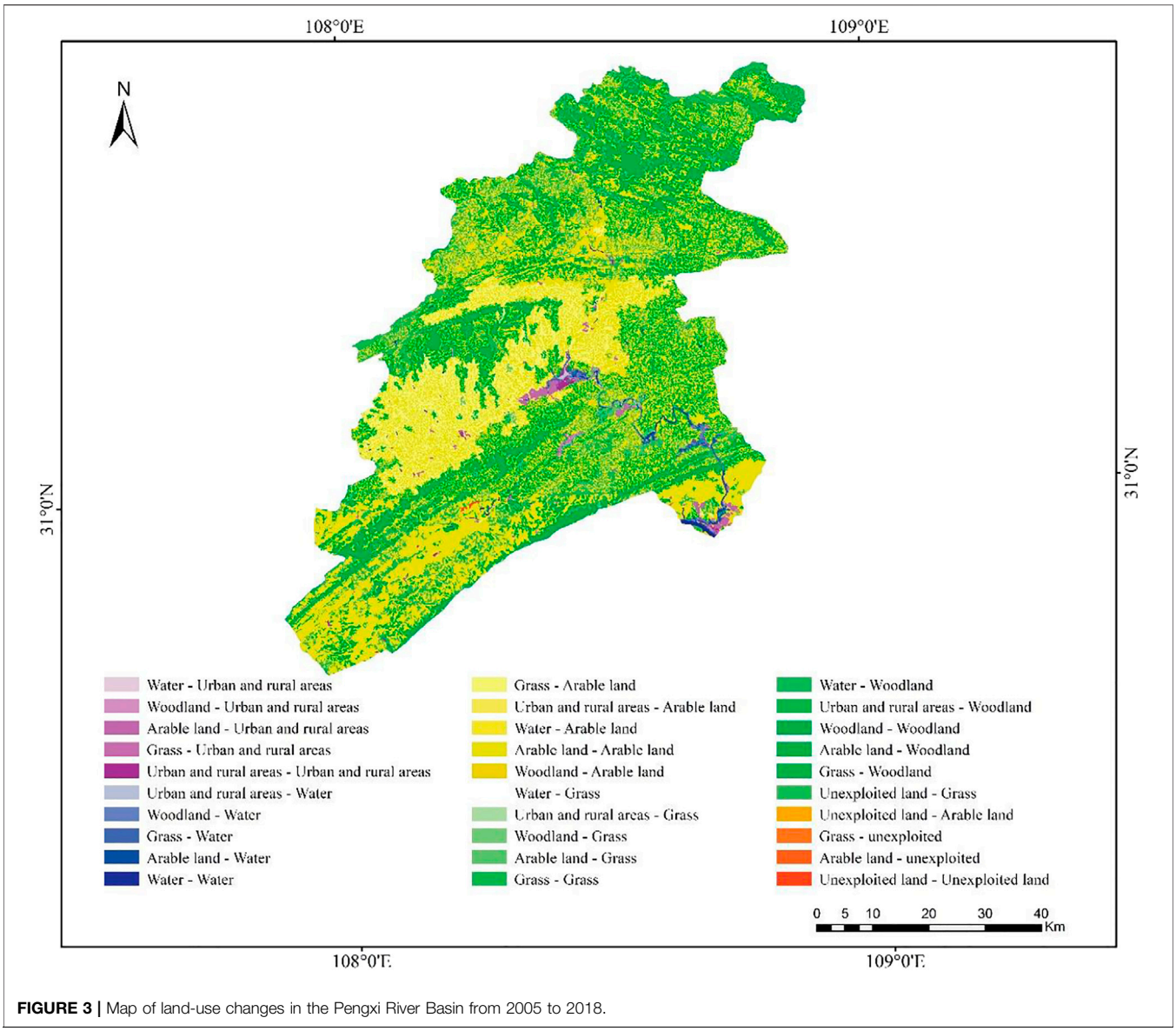


FIGURE 3 | Map of land-use changes in the Pengxi River Basin from 2005 to 2018.

potential evapotranspiration. The water available for vegetation growth was calculated based on the soil's sand, silt, clay, and organic matter contents. The plant evapotranspiration coefficient was calculated based on the leaf area index (LAI). The saturated water conductivity of the soil was calculated using the soil's sand and clay contents and the soil bulk density. The soil depth and DEM were extracted from the soil dataset of the Harmonized World Soil Database (HWSD) to calculate the topographic index.

2.3 Methods of Evaluation of Water Conservation Service Functions

The water conservation module of the InVEST model calculates the water yield based on the water cycle principle using parameters such as the average annual precipitation, surface and vegetation evapotranspiration, and soil and vegetation root depth. Then, it obtains the final amount of water conservation using the velocity coefficient, topographic index, and saturated soil water conductivity. The specific calculation formula is as follows:

$$Water_{Retention} = min\left(1, \frac{249}{Velocity}\right) \times min\left(1, \frac{0.9 \times TI}{3}\right) \times min\left(1, \frac{Ksat}{300}\right) \times Yield, \quad (1)$$

where *water retention* is the water retention depth (mm), *velocity* is the velocity coefficient, and *TI* is the dimensionless topographic index, which is calculated using **Eq. 2**. *Ksat* is the saturated soil water conductivity (cm/d), which is calculated using the Saxton

model and **Eq. 3** (Li et al., 2019). *Yield* is the water yield, which can be calculated using **Eq. 4**.

$$TI = lg\left(\frac{Grid_{number}}{Soil\ Depth \times Percent\ Slope}\right), \tag{2}$$

where $Grid_{number}$ is the number of grids in the study area, *Soil Depth* _is the soil depth (mm), and *Percent Slope* is the slope percentage.

$$\begin{aligned} Ksat &= 22.644 \times e^{power} \\ power &= -0.995 + 0.014sand \\ &+ \frac{\left(-376.988 + 3.323sand - 2.329clay + 0.113clay^2\right)}{\theta_s}, \end{aligned} \tag{3}$$

where *Ksat* is the saturated soil water conductivity (cm/d), *sand* and *clay* are the soil sand and clay contents (%), and θ_s is the saturated soil water content (%).

$$Y_{jx} = \left(1 - \frac{AET_{xj}}{P_x}\right) \times P_x, \tag{4}$$

where Y_{jx} is the annual water output, P_x is the average annual rainfall in grid unit X, and AET_{xj} is the annual average evapotranspiration in grid unit X in ecosystem type J (Wang et al., 2021), which can be calculated using **Eq. 5**.

$$\frac{AET_{xj}}{P_x} = \frac{1 + \omega_x R_{xj}}{1 + \omega_x R_{xj} + \frac{1}{R_{xj}}}, \tag{5}$$

where R_{xj} is the dimensionless dryness index of grid unit X in ecosystem type J, which is calculated based on the ratio of the potential evapotranspiration to rainfall and using **Eq. 6**. Ωx is the dimensionless ratio of the annual available water to rainfall for modified vegetation.

$$R_{xj} = \frac{k \times ET_0}{P_x}, k = min\left(1, \frac{LAI}{3}\right), \tag{6}$$

where k is the vegetation coefficient, which can be calculated using the LAI. ET_0 is the evapotranspiration when the flat ground is completely covered by certain short green vegetation and the soil is still sufficiently moist (mm/d) (Zhang et al., 2012). **Eq. 7** is used to calculate the evapotranspiration.

$$ET_0 = 0.0013 \times 0.408 \times RA \times \left(T_{avg} + 17\right) \times \left(TD - 0.0123P\right)^{0.76}, \tag{7}$$

where *RA* is the solar atmospheric top layer radiation ($MJ/m^2 \cdot d$), T_{avg} and *TD* are the mean and difference of the mean of the daily maximum and minimum temperature (°C), respectively, and P is the monthly precipitation (mm).

$$\omega_x = Z\frac{AWC_x}{P_x}, \tag{8}$$

where *Z* is the Zhang coefficient (constant) and its range is (1–10), which represents the seasonal characteristics of the precipitation. When the precipitation is mostly concentrated in winter, *Z* approaches 10. When the precipitation is relatively concentrated in summer or is evenly distributed in all seasons, *Z* approaches 1. According to the model test and the analysis of the precipitation data, the precipitation in the study area is mainly concentrated from May to October, of which the precipitation from June to August accounts for about 50%. In addition, by referring to the research results of other scholars regarding the Three Gorges Reservoir area (Xiao et al., 2015), it was found that

$$AWC_x = min\left(max\ Soil_{Depth_x}, Root_{Depth_x}\right) \times PAWC_x, \tag{9}$$

where $maxSoil\_Depth_x$ is the maximum soil depth, $Root\_Depth_x$ is the root depth, and $PAWC_x$ is the water available for vegetation growth, which is calculated using **Equation 10**:

$$\begin{aligned} PAWC_x &= 54.509 - 0.132sand - 0.003\left(sand\right)^2 \\ &- 0.055silt - 0.006\left(silt\right)^2 - 0.738clay \\ &+ 0.007\left(clay\right)^2 - 2.688OM + 0.501\left(OM\right)^2, \end{aligned} \tag{10}$$

where *sand*, *silt*, *clay*, and *OM* are the sand, silt, clay, and organic matter contents of the soil (%), respectively.

3 RESULTS

3.1 Analysis of Temporal and Spatial Characteristics of Regional Environmental Changes

3.1.1 Simulation Scenarios

After the completion of the Kaizhou water-level regulation dam in 2017, the Kaizhou water-level regulation dam and the Three Gorges Dam formed a nested operation to influence the area of water conservancy projects. The highest and lowest water levels of the Three Gorges Project are 175 m (hereinafter referred to as the high water level) and 145 m (hereinafter referred to as the low water level), based on which surface cover data were obtained for the high and low water levels at two time nodes in the study area.

Second, the ecosystem service functions of the water conservation were the result of the combined effect of the meteorological and hydrological factors. Therefore, for precipitation and temperature, seven simulation scenarios were designed based on the meteorological data from 1959 to 2019 for the study area. For precipitation, a wet year (P_H), normal year (P_N), and dry year (P_L) were defined. A high-temperature year (T_H) and low-temperature year (T_L) were defined for air temperature. In view of the dual effects of precipitation and temperature, we defined a high-temperature year with little rain (T_HP_L) and a low-temperature year with abundant rain (T_LP_H).

3.1.1.1 Land Cover Changes

In order to reveal the impacts of human activities on the land cover, land cover data for two time nodes were selected to analyze the direction and source of the land loss of each land-use type during the study period using an ecosystem type transfer matrix. Before and after the operation of the nested water conservancy projects began, the area of grassland converted to other land-use types was the largest, followed by cultivated land, and the amount

TABLE 2 | Dynamic change of land cover.

Land-use type	2005	2018	2005–2018	
	Area (km^2)	Area (km^2)	Area (km^2)	Dynamic attitude (%)
Paddy fields	671.90	662.68	–9.22	–0.11
Dryland	1576.05	1897.16	321.11	1.57
Forest land	427.87	459.92	32.05	0.58
Shrub land	421.84	408.17	–13.68	–0.25
Open forest land	672.75	695.37	22.62	0.26
Other woodland	127.00	114.18	–12.82	–0.78
High coverage grassland	79.03	108.27	29.24	2.85
Medium coverage grassland	957.07	511.19	445.88	–3.58
Low coverage grassland	8.12	7.55	–0.57	–0.54
Graff	4.87	38.57	33.70	53.28
Reservoir pits	5.37	12.80	7.42	10.63
Beaches	7.06	2.18	–4.88	–5.31
Urban land	8.29	15.13	6.85	6.35
Rural residential land	4.82	13.12	8.29	13.23
Land for industry, mining, and transportation	0.01	25.78	25.77	14684.10
Bare rock gravel fields	1.27	1.27	–0.003	–0.02

of dryland converted was the smallest. The area of cultivated land was the largest, followed by woodland. The area of grassland converted was the largest (**Figure 3**) and was 523.38 km^2. The converted areas were mainly medium and high coverage grassland, and the area of medium coverage grassland converted (509.03 km^2), accounted for 45.85% of the total area of grassland converted, while the area of high coverage grassland converted was only 12.71 km^2 (1.15%). More grassland with a medium coverage was converted into the dryland (403.79 km^2), accounting for 42.19% of the total area converted. The high coverage grassland was converted into the woodland (7.20 km^2), dryland (3.79 km^2), and sparse woodland (0.55 km^2). The 22.28% of the farmland was converted to the dryland, and the dryland was mainly converted to sparse woodland, paddy fields, and medium coverage grassland. The percentage of farmland converted was the largest (61.39%). Among them, the area of dryland converted was the largest (568.45 km^2). The second largest was woodland (20.84%), among which the area of sparse forest land converted was the largest (111.37 km^2), and it was mainly converted from the medium coverage grassland, dryland, and paddy fields.

The results revealed that from 2005 to 2018, the areas of the dryland, open woodland, high coverage grassland, canals and reservoirs pits, urban land, land for rural residents, and land for industry, mining, and transportation increased; the areas of the paddy fields, shrub land, forest land, other beaches and coverage, low coverage grassland, and the grass in the bare rock gravel land decreased. These changes were mainly due to the nested operation of the water conservancy projects and corresponded to the impacts of human activities.

The dynamic changes in the land cover are presented in **Table 2**. Before the water conservancy project nesting was implemented in 2005, the land cover was dominated by the dryland (31.69%), medium coverage grassland (19.24%), and sparse forest land (13.53%). In 2018, the dryland (38.15%), sparse forest land (13.98%), and paddy fields (13.32%) were the main land-use types. After the nested water conservancy projects were implemented, the areas of the dryland and medium coverage grassland changed significantly, and the area of the dryland increased by 321.11 km^2. The area of the medium coverage grassland decreased the most by 445.878 km^2.

From 2005 to 2018, the dynamic attitudes of the urban and rural areas, water area, cultivated land, and forest land were 23.98, 16.11, 1.07, and 0.13%, respectively. The dynamic attitude of the industrial, mining, and transportation land was the largest (14,684.10%). The dynamic attitude of the water channels was the second largest (53.28%). The dynamic attitudes of the grassland and bare gravel land were –3.07% and –0.02%, respectively. Due to the construction of the Three Gorges Dam, the total reservoir area of about 55.5 km^2 in the Kaizhou District was submerged, and the grace length of the 400 km reservoir was submerged. The urban pattern of the district was completely changed. The old city was basically submerged, and the population in the district affected by this change accounted for 16.62% of the total immigrant population in Chongqing (Zeng and Wang, 2017). The migration period (1993–2008) accelerated the urbanization process in the Kaixian County, and the dynamic attitudes of the urban and rural land and cultivated land in the study area were relatively large. In addition, the construction of new towns in the county promoted the development of the urbanization pattern of the Kaizhou District.

3.1.1.2 Precipitation Characteristics

The total rainfall in the wet years was 1716.4 mm, which mainly occurred from June to October. The annual rainfall in normal years was about 869.3 mm, which was 49.36% lower than that in wet years, and the average daily rainfall was 6.44 mm. The total rainfall in the dry years was only 759.7 mm, which was the lowest compared to the other years and was 55.77% lower than that in the wet years.

Precipitation was abundant in the low-temperature years, with the annual rainfall reaching 1180.7 mm. The annual total rainfall in the high-temperature years was the same as that in the normal years, which was 26.37% lower than in the low-temperature years.

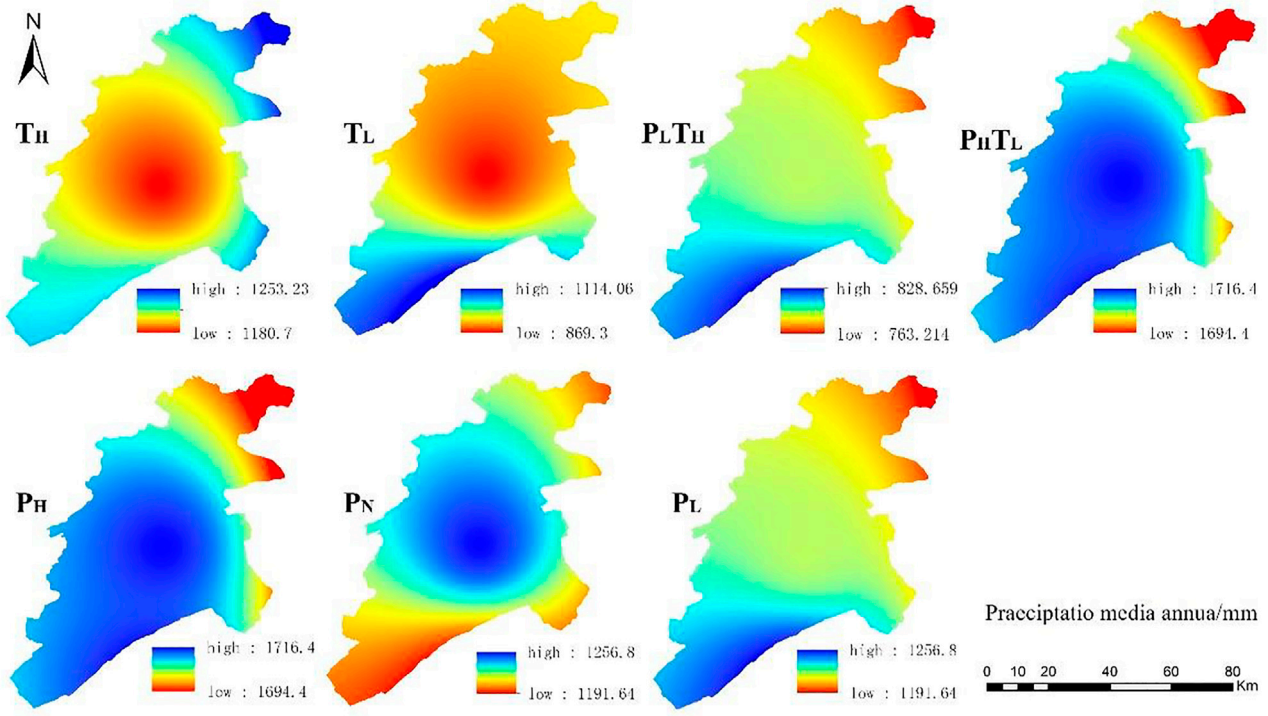


FIGURE 4 | Annual precipitation in the Pengxi River Basin under different scenarios.

The annual precipitation in Fengshui exhibited an increasing trend from the Mingzhong Township in the northeast to the southwest (**Figure 4**), and the rate of increase was rapid. The annual precipitation in the normal years and high-temperature years decreased from the central area to the Toudao River to the northeast and southwest, and the precipitation was lower in the southwest than that in the northeast. The distribution of the rainfall was roughly the same in the dry and wet years, that is, the precipitation decreased from southwest to northeast, but the annual precipitation was lower in the dry years. However, the annual precipitation in the low-temperature years decreased from the northeast to the central part of the study area and increased from the central part of the study area to the Kanto River in the southwest. Moreover, the precipitation was lower in the southwest than that in the northeast. The annual precipitation in the high temperature and low rainfall years increased slowly from northeast to southwest in the study area. In the low temperature and rainy years, the precipitation increased from northeast to southwest.

In general, the precipitation increased from northeast to southwest in the wet years, dry years, high temperature and low rainfall years, and low temperature and high rainfall years, with a rapid increase in wet years and a slow increase in dry years. The precipitation characteristics were the same in the low temperature and high rainfall years and high temperature and low rainfall years, and the precipitation characteristics were the same in the high temperature and low rainfall years and dry years. The annual precipitation in the normal years and high-temperature years decreased from the central China to the northeast and southwest. The precipitation in the low-temperature years increased from central China to the northeast and southwest, and the precipitation in the northeast was greater than that in the southwest. Although the spatial distribution characteristics of the precipitation were similar under the different scenarios, the precipitation itself was quite different in different scenarios.

3.1.1.3 Temperature Characteristics

The average daily temperatures in the wet years, normal years, and dry years were 17.97, 19.61, and 18.41°C, respectively. The annual average temperatures in the Pengxi River Basin in the high temperature and low-temperature years were 18.69 and 16.84°C, respectively. Under the two extreme scenarios, the trends of the temperature and precipitation were similar. The temperature initially gradually increased and then began to decrease in August. The precipitation exhibited an increasing–decreasing–increasing–decreasing trend. The difference in the potential evapotranspiration between the high-temperature years and low-temperature years in the study area is 250.51–305.53 mm/d (**Figure 5**). The average temperature in the high-temperature years was 19.4°C, and that in the low-temperature years was 17.42°C.

The results of the potential evapotranspiration calculation are shown in **Figure 5**. The distributions of the potential evapotranspiration in the high-temperature years and low temperature and high rainfall years were roughly the same, increasing from the central part of the area to the Toudao River in the axial direction. In contrast, in the normal years, it increased from the northeast and southwest to the central part of the area. In general, the distribution of the potential

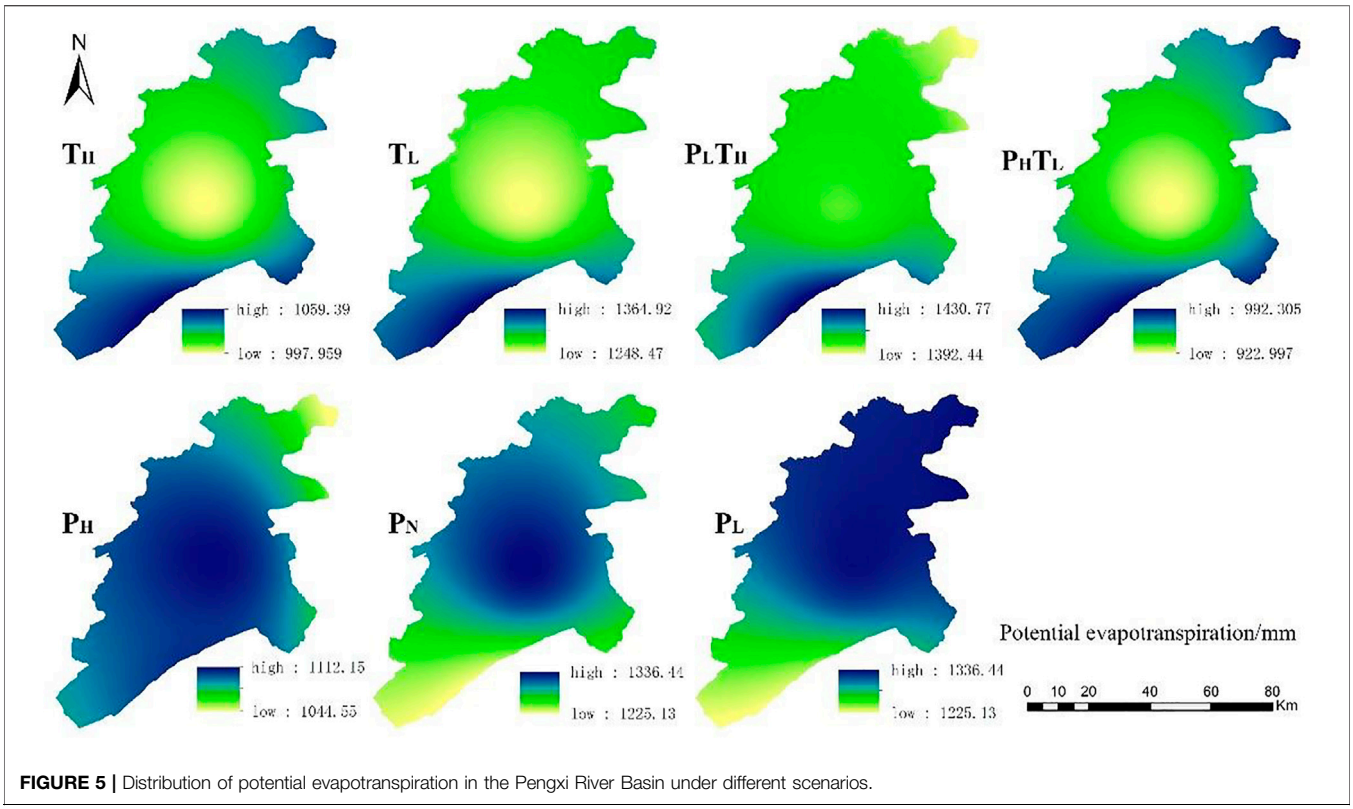


FIGURE 5 | Distribution of potential evapotranspiration in the Pengxi River Basin under different scenarios.

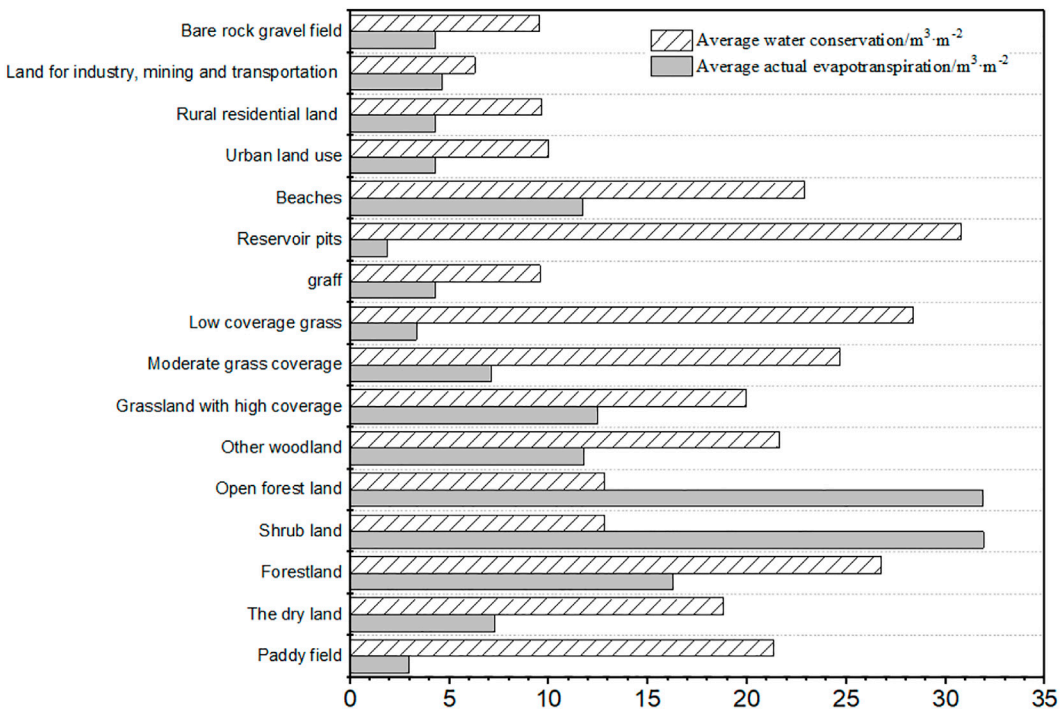


FIGURE 6 | Comparison of water conservation and actual evapotranspiration of the different types of ecosystems.

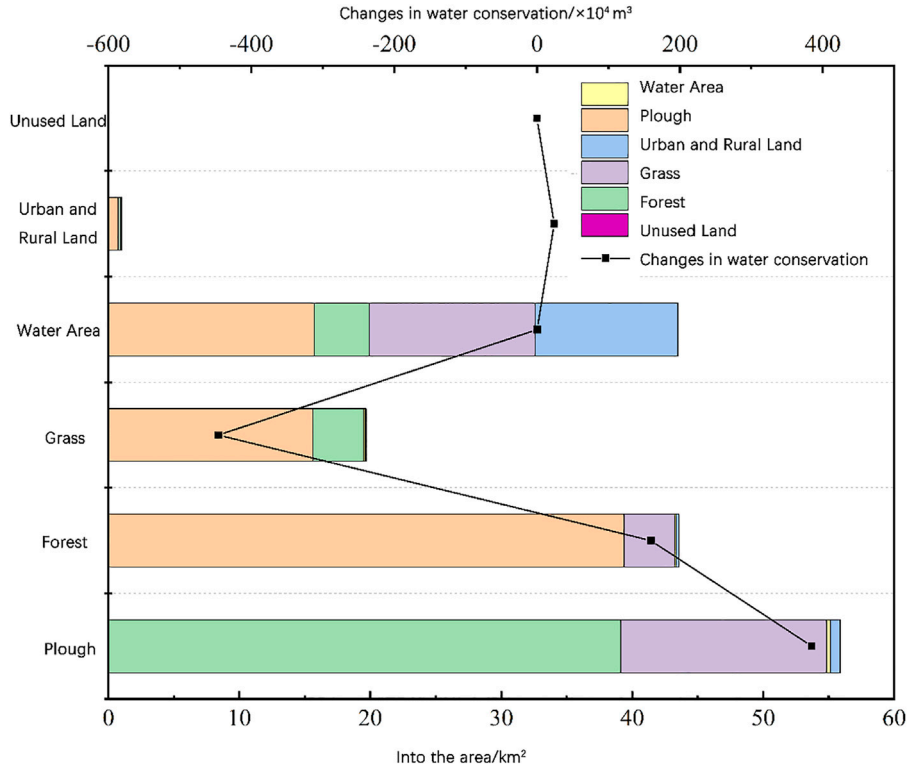


FIGURE 7 | Land-use changes and water conservation changes at different water levels in 2018.

TABLE 3 | Land-use changes and water conservation changes at different water levels in 2018.

Land-use type	Variable-area (km^2)	Low water-level water conservation (ten thousand m^3)	High water-level water conservation (ten thousand m^3)	Changes in water conservation (ten thousand m^3)
Arable land	−15.58	2523.42	2509.47	−13.95
Woodland	−4.01	7031.31	7017.31	−14.01
Grassland	−12.61	796.56	773.65	−22.90
Water	42.96	35.09	67.86	**32.77**
Urban and rural land	−10.88	38.30	30.55	−7.75
Unused land	0.12	0.87	1.17	0.30
Total	0	10425.55	10400.01	−25.54

evapotranspiration was basically consistent with the precipitation and air temperature distributions, which preliminarily indicates that the water conservation was strongly correlated with the evapotranspiration.

3.2 Analysis of Effects of Land Cover Changes on the Water Conservation Function

3.2.1 Effects of Different Types of Ecosystems on the Functional Characteristics of the Water Conservation

Under the same precipitation and temperature conditions, the water conservation and actual evapotranspiration of different surface cover types were compared, and it was found that the water conservation values of the open forest land, shrub land, dryland, and forest land were higher. Among them, the open forest had the highest water conservation ($19.5751 \times 10^6\ m^3$) (**Figure 6**), but the shrub ecosystem had the highest average water conservation, which was 94.10% higher than the average water conservation of the reservoirs and ponds. The average water conservation values of the paddy fields and dryland were low, but they increased by 56.38 and 288.30% compared with those of the reservoirs and pits. The total actual evapotranspiration was higher in the dryland, paddy fields, medium coverage grassland, and forest land. The bare rock and gravel land and rural residential land had the lowest average evapotranspiration, but the average evapotranspiration of the low coverage grassland and reservoirs and pools increased by 286.15 and 230.88%,

TABLE 4 | Average variations in water conservation and actual evapotranspiration in the Pengxi River Basin in 2005 and 2018 (10^4 m^3).

Land-use type	2005		2018		Variation	
	Water conservation	Actual evapotranspiration	Water conservation	Actual evapotranspiration	Water conservation	Actual evapotranspiration
Arable land	1020.13	6140.32	1203.04	6997.91	182.90	857.60
Woodland	3608.12	3956.17	3672.82	4076.83	64.70	120.66
Grassland	524.72	3426.49	339.25	2029.40	−185.47	−1397.09
Water	26.12	109.68	29.08	132.59	2.96	22.91
Urban and rural land	4.56	16.24	20.89	58.65	16.33	42.42
Unused land	0.70	1.84	0.61	1.83	−0.09	−0.01
Total	5184.35	13650.74	5265.68	13297.22	81.33	−353.51

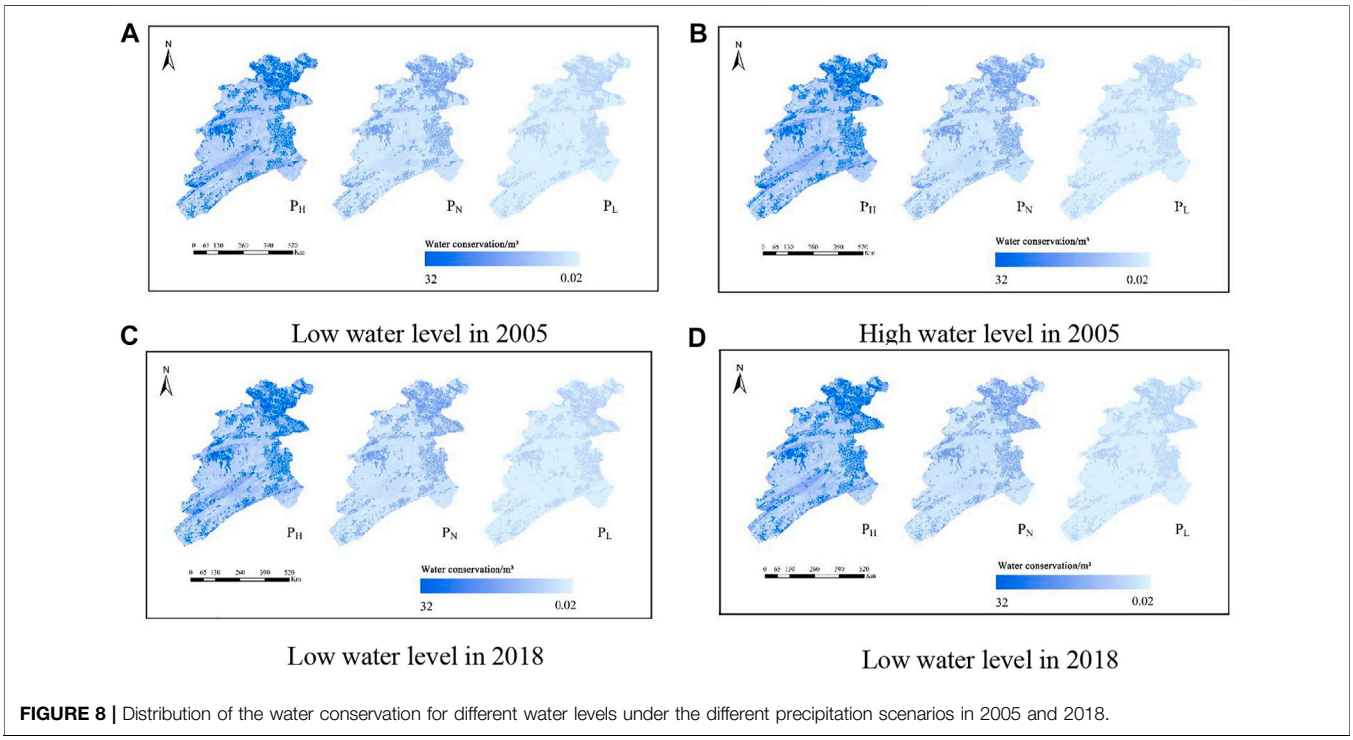


FIGURE 8 | Distribution of the water conservation for different water levels under the different precipitation scenarios in 2005 and 2018.

TABLE 5 | Response of water conservation function under different precipitation conditions (Unit: 10,000 m^3).

	Annual precipitation	Total water conservation	Total actual evapotranspiration
Wet years	1716.40	104255.54	13601.19
Dry years	795.74	49554.94	15169.89
Normal years	1256.86	16816.23	13081.99

respectively. Therefore, the water conservation function of the reservoirs was weak.

Overall, the water conservation exhibited a decreasing trend from southeast to northwest, and the water conservation was higher around the Liangma River, Xiaojiagou River, and Tuantan River. The actual evapotranspiration exhibited a decreasing trend from northwest to southeast, and the total evapotranspiration was higher around the Taoxi River, Qili River, and Baqu River.

3.2.2 Analysis of the Changes in the Water Conservation Function Caused by Changes in the Water Level

Under the same precipitation and temperature conditions (**Figure 7**), after the formal operation of the nested water conservancy projects began in 2018, except for the water area and unused land, the surface cover of the cultivated land, grassland, forest land, and urban and rural land at the low water level exhibited obvious decreasing trends compared with

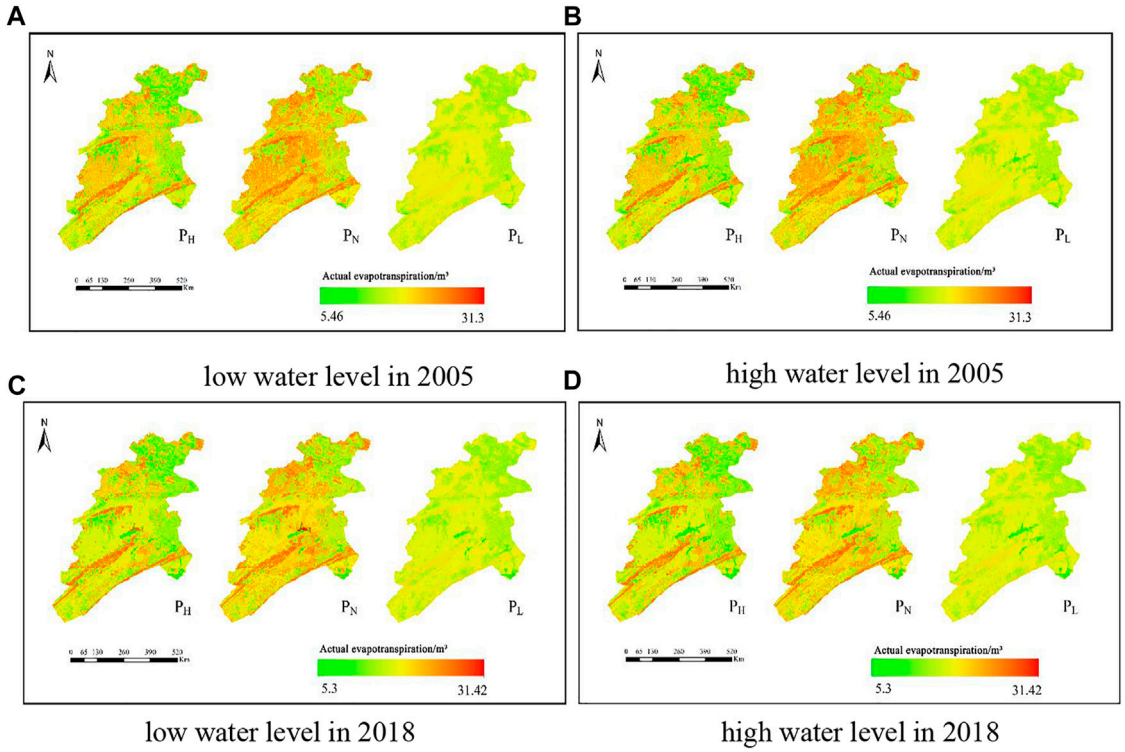


FIGURE 9 | Distribution of the total actual evapotranspiration for the different water levels under the different precipitation conditions in 2005 and 2018.

TABLE 6 | Response of water conservation function to evapotranspiration conditions.

	Cumulative daily mean air temperature (°C)	Total water conservation ($\times 10^4$ m³)	Total actual evapotranspiration ($\times 10^4$ m³)
High-temperature scenario	223.32	27355.28	13994.28
Low-temperature scenario	201.71	43715.36	12128.52

those at the high water level. The area of the high water level mainly increased in the vicinity of the Toudao River and its junction with Xiaojiagou. In the case of the high water level, the water conservation changed most obviously, with an average increase of 327,700 m^3 (**Table 3**).

The actual evapotranspiration increased the most in the cultivated land (8.5760×10^6 m^3), while the actual evapotranspiration of the grassland increased by -13.9709×10^6 m^3 (**Table 4**). The change in the water conservation of the cropland and grassland exhibited the opposite trend, and the total change in the water conservation of the cropland increased by 200% compared with that of the grassland, which was mainly caused by the construction of the water-level regulation dam and the immigration engineering in the reservoir area of the Three Gorges project. Through comprehensive comparison, it was found that the changes in the land cover had a certain impact on the water conservation and the actual evapotranspiration due to the difference in the evapotranspiration intensity and root depth in the different types of ecosystems. The change in the water level caused by the operation of the nested water conservancy projects was finally reflected in the change in the land cover, resulting in a reduction in the amount of water conservation (813,300 m^3) and the total actual evapotranspiration (3,535,100 m^3). Therefore, the construction of the Kaizhou water-level regulation dam and the Hanfeng Lake positively affected the water conservation in the region and improved the ecological service functions in the study area.

3.3 Analysis of Effects of Meteorological Changes on the Water Conservation Function

3.3.1 Analysis of Water Conservation Function Under Different Precipitation Conditions

Under different precipitation conditions (**Figure 8** and **Table 5**), the average water conservation depth in the wet years increased by 86.52 and 135.84 mm compared to the normal years and dry years, respectively, and the total water conservation increased by

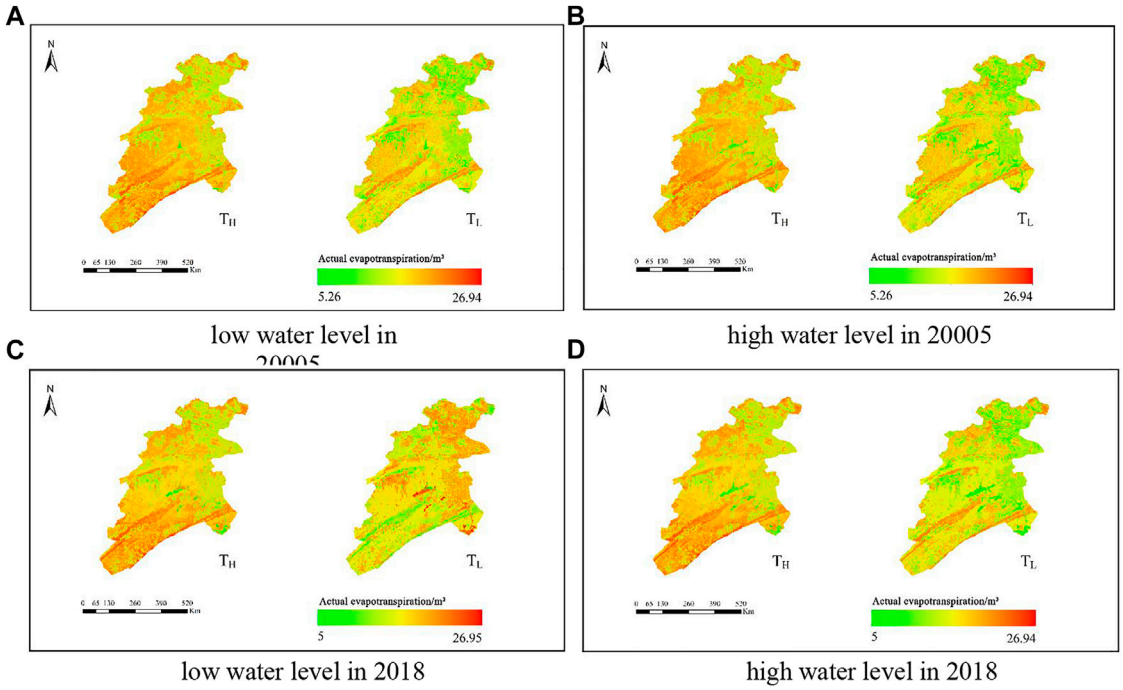


FIGURE 10 | Distributions of the total actual evapotranspiration for different water levels under different annual precipitation and evapotranspiration conditions in 2005 and 2018.

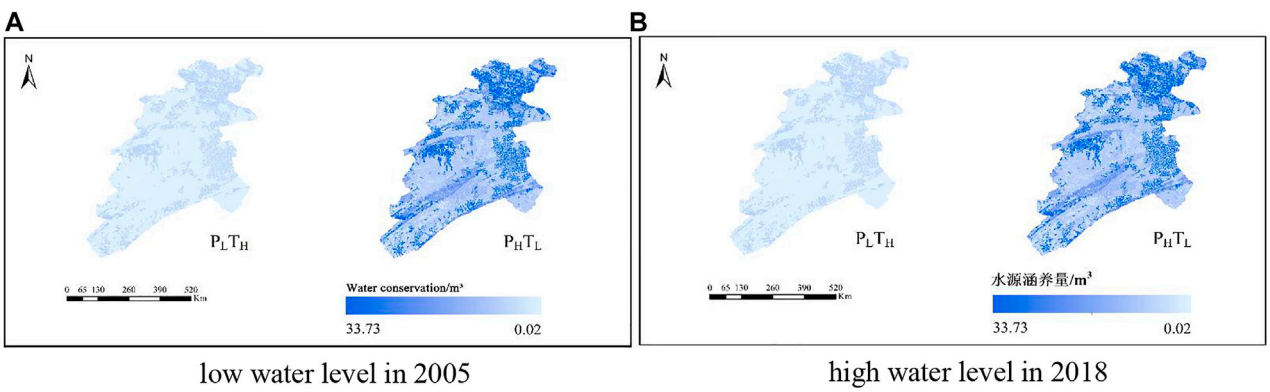


FIGURE 11 | Distribution of water conservation under extreme meteorological conditions for different water levels in 2005 and 2018.

32.62 and 17.55% under the low water level control. The effects of the different precipitation conditions on the water conservation were similar under the high water level and low water-level control. The results revealed that the influence of the precipitation conditions on water conservation was greater than that of the water level difference caused by the operation of the water conservancy projects. The precipitation conditions had a significant effect on the water conservation values of the different land cover types. The distribution of the water conservation was similar to that of the vegetation cover, that is, the high values are mainly distributed in the northeastern part of the Pengxi River Basin.

TABLE 7 | Change in the total water conservation under the influence of double extreme scenarios (unit: 10,000 m³).

Scenario	2005		2018	
	Low water	High water	Low water	High water
T_HP_L	1435.69	11171.44	1437.86	11111.18
T_LP_H	1481.47	11239.79	1483.68	11210.07

The average evapotranspiration and total actual evapotranspiration were the highest in the normal years, followed by the wet years, and were the lowest in the dry

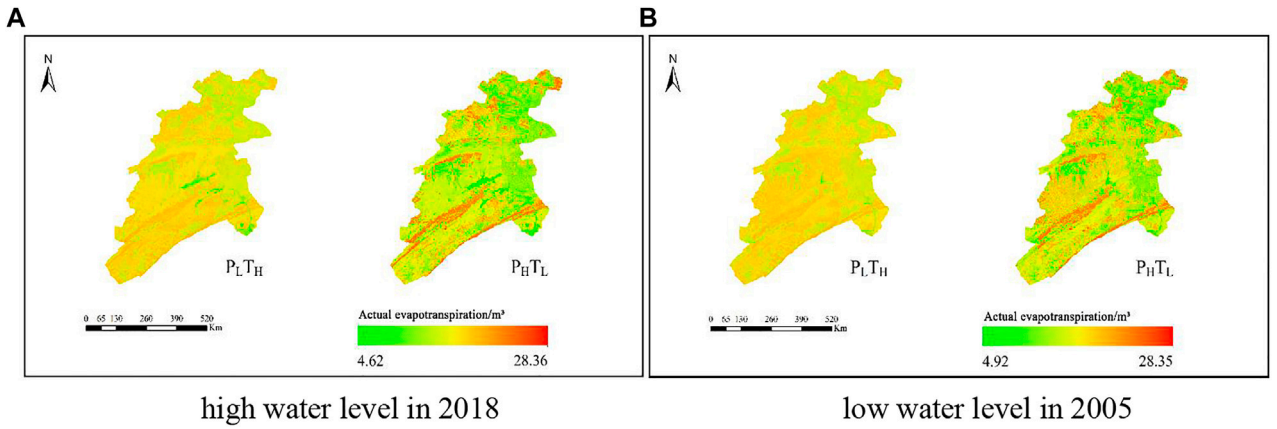


FIGURE 12 | Comparison of actual evapotranspiration for different water levels in 2005 and 2018.

TABLE 8 | Change in the total actual evapotranspiration under the influence of dual extreme scenarios (unit: 10,000 m³).

Scenario	2005		2018	
	Low water	High water	Low water	High water
T_HP_L	13673.06	12298.92	13608.55	12215.97
T_LP_H	13471.79	11860.21	13422.54	11795.54

years (**Figure 9**). Under the low water-level control, the average evapotranspiration in the dry years increased by 25.99 and 45.78 mm compared with the wet years and normal years, respectively, and the total actual evapotranspiration increased by 10.34 and 13.76%. Under the high water-level control, the average evapotranspiration in the normal years was 25.80 and 45.42 mm compared with that in the dry years and wet years, respectively. Spatially, the actual evapotranspiration was stronger in the eastern part of Xiaojiagou in the normal years. This was mainly because there were more paddy fields, grassland, and beaches in the east, so the potential evapotranspiration was relatively large and the change was obvious.

Overall, the average water conservation depth and total water conservation in the Pengxi River Basin were highest in the wet years, followed by the normal years, and were the lowest in the dry years. However, the average evapotranspiration and total actual evapotranspiration were the highest in the normal years, followed by the wet years, and were the lowest in the dry years.

3.3.2 Analysis of Effects of Temperature Conditions on the Water Conservation Function

Under the land cover conditions and low water-level control in 2005, the average water conservation depth under the low-temperature scenario was 25.27 mm higher (61.04%) than that under the high-temperature scenario. The average evapotranspiration under the high-temperature scenario was 29.72 mm higher (11.44%) than that under the low-temperature scenario. Under the land cover conditions and high water-level control in 2005, the average water conservation depth under the low-temperature scenario was 24.85 mm higher than that under the high-temperature scenario. The average evapotranspiration under the high-temperature scenario was 23.29 mm higher (9.19%) than that under the low-temperature scenario.

Under the land cover conditions and low water-level control in 2018 (**Table 6**), the annual average water conservation depth under the low-temperature scenario was 25.58 mm higher (63.56%) than that under the high-temperature scenario. The annual average evapotranspiration under the high-temperature scenario was 28.73 mm higher (11.13%) than that under the low-temperature scenario. Under the land cover conditions and high water-level control in 2018, the annual average water conservation depth under the low-temperature scenario was 25.66 mm (62.67%) higher than that under the high-temperature scenario. The annual average evapotranspiration under the high-temperature scenario was 29.82 mm higher (11.48%) than that under the low-temperature scenario.

If the water level and surface cover conditions are ignored, the water conservation level was about 60% higher under the low-temperature scenario than under the high-temperature scenario, and the total actual evapotranspiration was about 12–13% higher under the high-temperature scenario than under the low-temperature scenario. In particular, the temperature in the southwestern part of the Pengxi River Basin was higher under the high-temperature scenario, so the actual evapotranspiration in the southwestern part of the Pengxi River Basin was larger, which was in sharp contrast to the low-temperature scenario (**Figure 10**). Therefore, the air temperature was also one of the important factors controlling the water conservation function. Precipitation and temperature not only affected the total amount of water conservation, but also affected its spatial distribution.

3.3.3 Analysis of the Effects of Precipitation and Temperature on the Water Conservation Function

In the case of precipitation and temperature extremes, the changes in the water conservation function were significantly different (**Figure 11**; **Table 7**). Under the low-temperature and high rainfall scenarios, the water conservation reached the maximum value, and it was as high as 112.3979×10^6 m³ for

TABLE 9 | Model availability validation.

Year	Kaizhou			Yunyang			Modeled water production coefficient	Modeled water production modulus
	Precipitation depth (mm)	Water production coefficient	Water production modulus	Precipitation depth (mm)	Water production Coefficient	Water production modulus		
2005	1386.5	0.58	80.33	1196.8	0.48	57.35	0.43	46.872
2018	1146.5	0.48	55.58	1033.7	0.47	48.2	0.50	54.870

the high water-level control in 2018. In contrast, the lowest value occurred under the high temperature and low rainfall scenario, and it was as low as 14.3569×10^6 m³ under the low water-level control in 2005. There was an order of magnitude difference between them, that is, a difference of about 98×10^6 m³ (87.23%).

Similarly, under the two extreme evapotranspiration and precipitation scenarios for the study area, the average difference in the average water conservation was 9.6 m³ (**Figure 12**). From the perspective of the spatial distribution, the amount of water conservation was not only related to the type of ecosystem but also to the precipitation and air temperature distributions in the region (**Table 8**).

The maximum and minimum values of actual evapotranspiration increased in the Pengxi River Basin were 136,730,600 m³ and 117,953,300 m³, respectively, that is, a difference of 15.91%. The main reason for the high and low values is that the population and urban land in the study areas were mainly distributed near Qukou Town. When the water level changed, the main change in the land cover type was reflected in the water area and cultivated land, and the actual evapotranspiration per unit area of these two land cover types was higher.

4 DISCUSSION

The InVEST model was used to simulate the water-level changes in the study area and different meteorological conditions to obtain the amount of water supplied to the study area after subtracting the actual evapotranspiration from the precipitation and to calculate the water production coefficients and water production moduli of all of the scenarios. The results were compared with the water production modulus and water production coefficients for Kaizhou and Yunyang in Chongqing, which are located in the study area, obtained from the Chongqing Water Resources Bulletin (**Table 9**). The relative error of the validation model is 0.2, which shows that the model results are reliable.

The operation of the nested water conservation projects resulted in a positive trend in the water catchment's function. The operation of the water conservation projects has had a certain influence on the water conservation, but the meteorological conditions have had a greater impact on it. Moreover, the simulation results of the different precipitation and temperature scenarios show that the water conservation function responded more significantly to the precipitation conditions, which is consistent with the results of previous studies (Delphin et al., 2016; Dai and Wang, 2020). The depth of the water cover in this study is generally consistent with the precipitation trend. In addition, the water conservation function responded most significantly to the combined effect of the precipitation and temperature. In low temperature and high rainfall years, the water conservation was extremely high, especially under the high water-level control scenario after the operation of the nested water conservancy projects began, and the water conservation exceeded 100×10^6 m³. In contrast, the lowest values occurred in the high temperatures and low rainfall years, with water conservation of around 14×10^6 m³. The highest and lowest values are quantitatively different. Under the high-temperature scenario, the potential evapotranspiration was higher and the water conservation was lower. This is consistent with the findings of Chen et al. (2016) and Wang et al. (2021). Therefore, the influences of the complex environmental factors on the regional water conservation service function deserve further attention.

Climate change, management policies, and other factors have a certain impact on the function and value of ecological services (Luo et al., 2019; Marques et al., 2021). Relevant research results have shown that the terrain also has a certain impact on the evapotranspiration conditions and water conservation. Water conservation increases as the slope and topographic relief increase, and it fluctuates as the topographic position index increases (Xu et al., 2020). The topographic gradient significantly affects the water supply (Gao et al., 2016; Zhou et al., 2021). Hu et al. (2021) found that the correlation between the heterogeneity of the landscape and water conservation is higher than that between climate factors and water conservation (Hu et al., 2021). The impact of the landscape pattern on water-related ecosystem services is mainly reflected in the change in the intensity, but the dominant relationship does not change (Li et al., 2021b). Therefore, the influences of terrain and landscape heterogeneity on water conservation will be considered further in future studies.

5 CONCLUSION

The factors influencing the water conservation service function are complex and differ significantly in different situations. In the study area, there was a positive correlation between the land cover changes and water conservation changes. Among the land cover types, the forest ecosystem contributed the most to the water conservation, with an average contribution rate of 68%. The Pengxi River Basin is an important water conservation area in

northeastern Chongqing and is an important component of the Yangtze River ecological corridor.

Under the operation of nested water conservancy projects, the water conservation function has exhibited a positive development trend. The operation of the water conservancy project has had a certain effect on the water conservation function, but the meteorological conditions have had a greater effect. A comprehensive comparison of the water conservation function under seven simulated scenarios revealed that the changes in the precipitation and air temperature were the leading factors causing the changes in the water conservation function. However, there were differences in the mechanisms by which these two factors affected the water conservation function. Comparatively, precipitation had the greatest influence on the water conservation function.

DATA AVAILABILITY STATEMENT

The data analyzed in this study is subjected to the following licenses/restrictions: part of the dataset comes from the weather bureau. Data can only be used with the approval of the weather bureau. Requests to access these datasets should be directed to XD|, dxn98mry@sohu.com.

AUTHOR CONTRIBUTIONS

CZ and RL contributed to the conception and design of the study. YM, RL, and XD organized the database. WQ and YM performed the statistical analysis and created the figures. CZ and WQ wrote the first draft of the manuscript. CZ, WQ, YM, and BY wrote sections of the manuscript. All of the authors contributed to the revision of the manuscript and have read and approved the submitted version.

FUNDING

This research was financially supported by the Chongqing Natural Science Foundation Project (cstc2019jcyj-msxmX0227), the Chongqing Normal University Fund Project (No.19XL13011), the Chongqing Social Science Planning Project (2019QNGL23), the Belt and Road Special Foundation of the State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering (2021490211), the Chongqing University Innovation Research Group funding (No.CXQT20015), and the Science and Technology Research Project of the Chongqing Education Commission (KJQN201900524).

ACKNOWLEDGMENTS

The authors thank the Southwest Mountain Center of the National Earth System Science Data Center for providing the surface cover data and the Chongqing Meteorological Bureau for providing the precipitation and temperature data. We thank LetPub (www.letpub.com) for its linguistic assistance during the preparation of this article.

REFERENCES

Almeida, B., and Cabral, P. (2021). Water Yield Modelling, Sensitivity Analysis and Validation: A Study for Portugal. *ISPRS Int. J. Geo-Inf.* 10 (8), 494. doi:10.3390/ijgi10080494

Aneseyee, A. B., Soromessa, T., and Elias, E. (2022). Evaluation of Water Provision Ecosystem Services Associated with Land Use/Cover and Climate Variability in the Winike Watershed, Omo Gibe Basin of Ethiopia. *Environ. Manag.* 69, 367–383. doi:10.1007/s00267-021-01573-9

Benra, F., De Frutos, A., Gaglio, M., Álvarez-Garretón, C., Felipe-Lucia, M., and Bonn, A. (2021). Mapping Water Ecosystem Services: Evaluating InVEST Model Predictions in Data Scarce Regions. *Environ. Model. Softw.* 138, 104982. doi:10.1016/j.envsoft.2021.104982

Capon, S. J., Leigh, C., Hadwen, W. L., George, A., McMahon, J. M., Linke, S., et al. (2018). Transforming Environmental Water Management to Adapt to a Changing Climate. *Front. Environ. Sci.* 6, 80. doi:10.3389/fenvs.2018.00080

Chen, S.-S., Liu, K., and Bao, Y.-B. (2016). Spatial Pattern and Influencing Factors of Water Conservation Service Function in Shangluo City. *Sci. Geogr. Sin.* 2016 (10, 1546–1554). doi:10.13249/j.cnki.sgs.2016.10.012

Chen, W.-X., Zhao, X.-L., Zhong, M.-X., Li, J., and Zeng, J. (2022). Spatiotemporal Evolution Patterns of Ecosystem Health in the Middle Reaches of the Yangtze River Urban Agglomerations. *Acta Ecol. Sin.* 2022 (01, 2–12). doi:10.5846/stxb202012093142

Cheng, H., Wu, S.-J., Wang, X.-X., Jiang, Y., Chen, C.-D., Wang, Y., et al. (2015). Research Progress on the Effects of the Three Gorges Reservoir on the Ecological Environment. *Chin. J. Eco-Agriculture* 23 (02), 127–140. doi:10.13930/j.cnki.cjea.140734

Cheng, L., Opperman, J. J., Tickner, D., Speed, R., Guo, Q., and Chen, D. (2018). Managing the Three Gorges Dam to Implement Environmental Flows in the Yangtze River. *Front. Environ. Sci.* 6, 64. doi:10.3389/fenvs.2018.00064

Dai, E.-F., and Wang, Y.-H. (2020). Spatial Heterogeneity and Driving Mechanisms of Water Yield Service in the Hengduan Mountain Region. *Acta Geogr. Sin.* 75 (3), 607–619. doi:10.11821/dlxb202003012

Delphin, S., Escobedo, F. J., Abd-Elrahman, A., and Cropper, W. P. (2016). Urbanization as a Land Use Change Driver of Forest Ecosystem Services. *Land Use Policy* 54, 188–199. doi:10.1016/j.landusepol.2016.02.006

Gao, H.-J., Han, H.-Q., Yu, H.-Y., and Han, M.-R. (2016). Distribution Characteristic of Important Ecosystem Services in Terrain Gradient in Wujiang River Basin. *Ecol. Sci.* 2016 (05), 154–159. doi:10.14108/j.cnki.1008-8873.2016.05.021

Gong, X.-J., Wang, X.-F., Liu, T.-T., Yuan, X.-Z., Kong, W.-W., and Liu, H. (2020). Spatial-temporal Characteristics of Carbon,nitrogen and Phosphorus and Eutrophication Assessment in a Typical River of Three Gorges Reservoir under the Development of Field Towns. *J. Lake Sci.* 32 (01), 111–123. doi:10.18307/2020.0111

He, X.-B., and Bao, Y.-H. (2019). Research Advances on Soil Erosion and Ecological Restoration in the Riparian Zone of the Three Gorges Reservoir. *Sci. Soil Water Conservation* 17 (04), 160–168. doi:10.16843/j.sswc.2019.04.019

Hu, C.-H. (2019). Analysis on Sediment Scouring and Silting Variation of Three Gorges Reservoir since 175 M Trial Impoundment for Past Ten Years. *Water Resour. Hydropower Eng.* 50 (08), 18–26. doi:10.13928/j.cnki.wrahe.2019.08.003

Hu, W., Li, G., and Li, Z. (2021). Spatial and Temporal Evolution Characteristics of the Water Conservation Function and its Driving Factors in Regional Lake Wetlands—Two Types of Homogeneous Lakes as Examples. *Ecol. Indic.* 130, 1470. doi:10.1016/j.ecolind.2021.108069

Li, J., Zhou, K., Xie, B., and Xiao, J. (2021). Impact of Landscape Pattern Change on Water-Related Ecosystem Services: Comprehensive Analysis Based on Heterogeneity Perspective. *Ecol. Indic.* 133, 108372. doi:10.1016/j.ecolind.2021.108372

Li, M.-Y., Liu, T.-X., Luo, Y.-Y., Duan, L.-M., Zhang, J.-Y., and Zhou, Y.-J. (2019). Pedo-Transfer Function and Remote-Sensing-Based Inversion Saturated Hydraulic Conductivity of Surface Soil Layer in Xilin River Basin. *Acta Pedol. Sin.* 56 (1), 90–100. doi:10.11766/trxb201803200104

Li, Q.-Q., and Hao, F.-J. (2019). Deformation Trend Judgment and Prediction of Accumulated Layer Landslide in Three Gorges Reservoir Area. *Water Resour. Hydropower Eng.* 50 (05), 228–233. doi:10.13928/j.cnki.wrahe.2019.05.027

Li, W.-J., Wang, S., Li, Q., Wu, T.-L., and Zhao, X.-Y. (2021a). The Impacts of Climate and Land Use Changes on Water Yield in the Beisan River Basin. *J. East China Normal Univ. Sci.* 2021 (04), 99–108. doi:10.3969/j.issn.1000-5641.2021.04.012

Li, W.-J., Zhu, K., Ran, Y.-G., Ran, J.-J., and Huang, P. (2021b). Effects of Land-Use Types and Off-Season Water-Level Fluctuation on Component Characteristics of Soil Active Organic Carbon. *J. Soil Water Conservation* 35 (02), 178–183. doi:10.13870/j.cnki.stbcxb.2021.02.024

Lu, B. (2021). A Review of the Literature on Ecological Restoration of the Three Gorges Reservoir Area's Subsidence Zone. *Technol. Innovation Appl.* 11 (19), 1–8.

Lu, Z.-X., Feng, Q., Xiao, S., Xie, J., Zou, S., Yang, Q., et al. (2021). The Impacts of the Ecological Water Diversion Project on the Ecology-Hydrology-Economy Nexus in the Lower Reaches in an Inland River Basin. *Resour. Conservation Recycl.* 164, 105154. doi:10.1016/j.resconrec.2020.105154

Luo, Y., Lu, Y., Fu, B., Zhang, Q., Li, T., Hu, W., et al. (2019). Half Century Change of Interactions Among Ecosystem Services Driven by Ecological Restoration: Quantification and Policy Implications at a Watershed Scale in the Chinese Loess Plateau. *Sci. Total Environ.* 651 (Pt 2), 2546–2557. doi:10.1016/j.scitotenv.2018.10.116

Marques, S. M., Campos, F. S., David, J., and Cabral, P. (2021). Modelling Sediment Retention Services and Soil Erosion Changes in Portugal: A Spatio-Temporal Approach. *ISPRS Int. J. Geo-Inf.* 10 (4), 262. doi:10.3390/ijgi10040262

Pan, Y., Zheng, H., Yi, Q., and Li, R. (2021). The Change and Driving Factors of Ecosystem Service Bundles: A Case Study of Daqing River Basin. *Acta Ecologica Sinica.* 13, 5204–5213. doi:10.5846/stxb202103100650

Raji, S. A., Odunuga, S., and Fasona, M. (2021). Quantifying Ecosystem Service Interactions to Support Environmental Restoration in a Tropical Semi-arid Basin. *Acta Geophys.* 69 (5), 1813–1841. doi:10.1007/s11600-021-00644-z

Ren, M., and Mao, D.-H. (2021). Supply and Demand Analysis and Service Flow Research of Water Production Service in Lianshui River Basin. *Ecol. Sci.* 40 (02), 186–195. doi:10.14108/j.cnki.1008-8873.2021.02.023

Shen, Z.-F., Zhang, K.-J., Xia, X., and Hu, L. (2021). Bibliometric Analysis of the Current Situation and Hot Research Topics on the Water Level Fluctuation Zone (WLFZ) of Three Gorges Reservoir. *J. Hydroecology* 42 (01), 26–34. doi:10.15928/j.1674-3075.202006280187

Shrestha, M., Piman, T., and Grünbühel, C. (2021). Prioritizing Key Biodiversity Areas for Conservation Based on Threats and Ecosystem Services Using Participatory and GIS-Based Modeling in Chindwin River Basin, Myanmar. *Ecosyst. Serv.* 48, 101244. doi:10.1016/j.ecoser.2021.101244

Wang, B.-S., Chen, H.-X., Dong, Z., Zhu, W., Qiu, Q.-Y., and Tang, L.-N. (2020). Impact of Land Use Change on the Water Conservation Service of Ecosystems in the Urban Agglomeration of the Golden Triangle of Southern Fujian, China, in 2030. *Acta Ecol. Sin.* 40 (02), 484–498. doi:10.5846/stxb201902180286

Wang, R., Peng, Q., Zhang, W., Liu, C., and Zhou, L. (2022). Ecohydrological Service Characteristics of Qilian Mountain Ecosystem in the Next 30 Years Based on Scenario Simulation. *Sustainability* 14, 1819. doi:10.3390/su14031819

Wang, X.-F., Fu, X.-X., Chu, B.-Y., Li, Y., Yan, Y., and Feng, X.-M. (2021). Spatio-temporal Variation of Water Yield and its Driving Factors in Qinling Mountains Barrier Region. *J. Nat. Resour.* 36 (10), 2507–2521. doi:10.31497/zrzyxb.20211005

Xiao, Y., Xiao, Q., Ouyang, Z., and Maomao, Q. (2015). Assessing Changes in Water Flow Regulation in Chongqing Region, China. . *Environ. Monit. Assess.* 187. doi:10.1007/s10661-015-4370-4

Xu, C.-X., Gong, J., Li, Y., Yan, L.-L., and Gao, B.-L. (2020). Spatial Distribution Characteristics of Typical Ecosystem Services Based on Terrain Gradients of Bailongjiang Watershed in Gansu. *Acta Ecol. Sin.* 40 (13), 4291–4301. doi:10.5846/stxb201911152447

Zeng, W., and Wang, L.-L. (2017). Research on Migrant Community Optimization in the Three Gorges from the Perspective of Social Network. *Chongqing Archit.* 2017 (07), 5–10. doi:10.3969/j.issn.1671-9107.2017.07.005

Zhang, H.-F., Ouyang, Z.-Y., and Zheng, H. (2007). Spatial Scale Characteristics of Ecosystem Services. *Chin. J. Ecol.* 26 (09), 1432–1437. doi:10.13292/j.1000-4890.2007.0247

Zhang, X.-L., Xiong, L.-H., Lin, L., and Long, H.-F. (2012). Application of Five Potential Evapotranspiration Equations in Hanjiang Basin. *Arid. Land Geogr.* 2012 (02), 229–237. doi:10.13826/j.cnki.cn65-1103/x.2012.02.006

Zhao, Y.-R., Zhou, J.-J., Lei, L., Xiang, J., Huang, M.-H., Feng, W., et al. (2019). Identification of Drivers for Water Yield in the Upstream of Shiyang River Based on InVEST Model. *Chin. J. Ecol.* 38 (12), 3789–3799. doi:10.13292/j.1000-4890.201912.017

Zhou, Z.-M., Yin, H.-R., Zhu, W.-B., and Ye, L.-P. (2021). Influence of Topography on Water Supply in Qihe Basin of Taihang Mountains. *J. Henan Univ. Sci.* 2021 (04), 379–388. doi:10.15991/j.cnki.411100.2021.04.001

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Environmental Science

ORIGINAL RESEARCH
published: 18 July 2022
doi: 10.3389/fenvs.2022.907769

Characteristics and Pollution Contribution of the Internal Nitrogen Release From the Sediments in the Dahekou Reservoir in Inner Mongolia

Junping Lu, Tingxi Liu*, Xiao Hong Shi, Biao Sun and Shengnan Zhao*

Water Resources Protection and Utilization Key Laboratory, Water Conservancy and Civil Engineering College, Inner Mongolia Agricultural University, Hohhot, China

To clarify the influence of the changes in the overlying water environment on the internal nitrogen release from reservoir sediments in different seasons, the quantitative linear relationship between the intensity of the nitrogen release from the sediment and the environmental factors of the overlying water was established, and their contribution rate to the nitrogen pollution of the reservoir during different storage periods was investigated. In this study, the sediment samples were collected from the Dahekou Reservoir in the Xilingol League, and the orthogonal simulation experiments were conducted in the laboratory. The mathematical model, which was established using multiple linear regression methods, revealed the following. The order of the significance of the influences of the environmental factors on the nitrogen release from the sediments in the Dahekou Reservoir is water temperature (T) > dissolved oxygen (DO) > pH value > hydrodynamic force (K). The total nitrogen release flux from the sediments in the Dahekou Reservoir was 14.278 t/a in 2018, accounting for 27.91% of the total nitrogen (TN) pollution load input during the same period. In particular, in winter, the contribution rate of the nitrogen released from the sediments reached the highest level (57.06–63.26%), which was significantly higher than the river's contribution to the total nitrogen pollution load of the reservoir. The nitrogen released from the sediments became the main source of nitrogen nutrients in the reservoir in the ice-sealed period.

Keywords: northern cold area, Dahekou Reservoir, endogenous pollution, sediment, nitrogen release model, contribution rate

OPEN ACCESS

Edited by:
Zhenzhong Zeng,
Southern University of Science and Technology, China

Reviewed by:
Changhui Wang,
Nanjing Institute of Geography and Limnology (CAS), China
Guoqing Wang,
Nanjing Hydraulic Research Institute, China

****Correspondence:***
Junping Lu
junping516@163.com
Tingxi Liu
txliu1996@163.com

Specialty section:
This article was submitted to Freshwater Science, a section of the journal Frontiers in Environmental Science

Received: *30 March 2022*
Accepted: *13 June 2022*
Published: *18 July 2022*

Citation:
Lu J, Liu T, Shi XH, Sun B and Zhao S (2022) Characteristics and Pollution Contribution of the Internal Nitrogen Release From the Sediments in the Dahekou Reservoir in Inner Mongolia. Front. Environ. Sci. 10:907769. doi: 10.3389/fenvs.2022.907769

1 INTRODUCTION

Eutrophication of water has been a concern since the early 20th century. The eutrophication of water bodies was first studied in the Great Lakes region of North America. According to the survey results of Davis (1964) for Lake Erie in North America, the number of phytoplankton's in the lake continued to increase, reaching the maximum in spring and autumn and lasting for an increasingly long time. Subsequently, Beeton's (1965) research confirmed that the acceleration of the Lake Erie eutrophication was caused by human activities. In addition to the United States, Japan (Okino and Kato, 1987), Germany (Hartman and Nümann, 1977), Sweden (Gelin, 1975), Australia (Hammer et al., 1973), and other countries have also conducted studies on lake eutrophication. Since eutrophication was first discovered in the 1930s, 30–40% of the world's lakes and reservoirs

have been affected by eutrophication to varying degrees. Taking effective measures to control the input of exogenous nitrogen and phosphorus nutrients into the upper reaches of reservoirs has become a common measure for the prevention and control of water eutrophication. However, the internal pollution caused by the release of nitrogen and phosphorus from sediments has been neglected. In this study, the term sediment generally refers to river and lake sediments, which is a common accumulation of particles. A variety of pollutants and numerous types of pollutant degradation, adsorption, desorption, and other biological and ecological effects lead to a variety of pollutant and natural material interactions after the combined effects of these factors have prevailed for a long period of time. Throughout the entire water system, material circulation serves the important roles of confluence and source (Yang et al., 2018). Previous studies have shown that the migration and transformation of nutrients at the mud-water interface are affected and restricted by various environmental factors such as water temperature (Zhong et al., 2008), pH (Bocrs and Istvanovics, 1991), water disturbance (Pang et al., 2008; Yan et al., 2008), and dissolved oxygen (Zhang et al., 2012). The released form, release time, and released quantity of the nutrient salts in the sediments have been found to be random (Hartman and Nümann, 1977; Okino and Kato, 1987; Xue and Lu, 2015). Under the gradual and effective control of the exogenous input, the release of nutrients from sediments becomes an important cause of the deterioration and eutrophication of the lake water quality (Mayer et al., 2005). Therefore, the secondary pollution caused by the internal release of nutrients from sediments has attracted the attention of many scholars (Ni and Wang, 2015; Gu et al., 2016). In the 1970s, Schindler (1974) conducted long-term large-scale experiments on a lake, and Schelske et al. (1974) investigated and studied the phosphorus and silicon dioxide in Lake Michigan, both of which revealed that phosphorus was the main limiting factor in the eutrophication of freshwater lakes. Therefore, phosphorus load reduction has become the main strategy for lake management in North America and Europe. Subsequently, research on lakes Apopka, George, and Okeechobee in the United States, East Lake in China, and Xiupu Lake in Japan have confirmed that using only this method of controlling the phosphorus input fails to control the eutrophication of the water body (Conley et al., 2002). The contribution of nitrogen to water eutrophication has increased significant interest among scientists, and a large number of studies have confirmed that nitrogen is another key limiting factor in the water eutrophication of freshwater lakes (Han et al., 2014; Shan et al., 2015). At present, most scholars have mainly focused on the influences of the lake water body sediment nitrogen composition, occurrence forms, spatial distribution, diffusion flux, and interface structure on the contribution to the water environment (Pennuto et al., 2014; Yang et al., 2017). Conversely, little research has been conducted on the nitrogen release from sediments, which is affected by changes in the overlying water environment. Most previous studies were conducted in the laboratory under static conditions to simulate the impact of changes in a single environmental factor on the release characteristics of the internal nitrogen in sediments, and few studies have investigated the impacts of the synergistic effects of various environmental factors on the release characteristics of the internal nitrogen in sediments. In addition, most studies on sediment nitrogen nutrient release are still in the qualitative stage, so it is urgent to further analyze and quantitatively discuss and estimate the sediment internal nitrogen release load and the pollution contribution rate to the overlying water body. Studies of the contribution of the exogenous input and internal release under the influence of seasonal changes, river runoff, and the water quality upstream of the reservoir will have a significant impact on the formulation of pollution prevention plans and measures for the reservoir.

The Dahekou Reservoir is in a cold and dry area with strong sand movement. The water quality exhibits a eutrophication level that has increased year by year. During winter and spring replacement and the freezing and thawing of the reservoir, the water temperature stratification and inversion influence the release of the nitrogen and nutrient salts in the sediments. Wind and reservoir water drainage disturb the reservoir sediments at the bottom, resulting in sediment suspension and nitrogen redistribution at the interface between the sediments and water (Xiang et al., 2015). Some nutrients can be released from sediments into the upper water, increasing the nutrient load (Holdren and Armstrong, 1980). Therefore, the temperature, pH, dissolved oxygen, hydraulic disturbance, and sediment mathematical model of the internal nitrogen release strength (lake) were considered when simulating the reservoir conditions at the soil interface under environmental changes to estimate the deposited silt sediment nitrogen release flux, to control the water source pollution, to prevent eutrophication, and to ensure the reservoir and downstream water supply function.

2 MATERIALS AND METHODS

2.1 Survey of Study Areas

The Dahekou Reservoir (42°13′19.17″N, 116°38′4.00″E) is located on the main stream of the Luanhe River in Duolun County, Xilingol League, Inner Mongolia. It was built in August 1995, and it has a maximum water depth of 18.1 m, a designed reservoir capacity of 26 million m^3, an area of 8684.54 km^2 above the dam site, an average monthly area of 2.14–2.87 km^2, and a water level of 1213.39–1215.13 m. The maximum water temperature for many years is 23.3°C, and the average water temperature under the ice during the ice period is 0.4°C. The water level, area, and depth of the reservoir vary with the storage and discharge of the Datang Duolun Hydropower Station at the mouth of the dam. It is a medium-sized narrow reservoir, and water supply and power generation are its main purposes, as well as comprehensive utilization of flood control, agricultural irrigation, and aquaculture. It is mainly recharged by the surface runoff of the Tuligen River in the east and Luanhe River in the west. For many years, due to the influence of sewage discharge from the sewage treatment plant in the upper reaches of Duolun County and the fishery culture in the reservoir, the organic matter (e.g., humus) in the sediments of the reservoir rot, and the nitrogen and phosphorus contents increase year by year. The eutrophication of

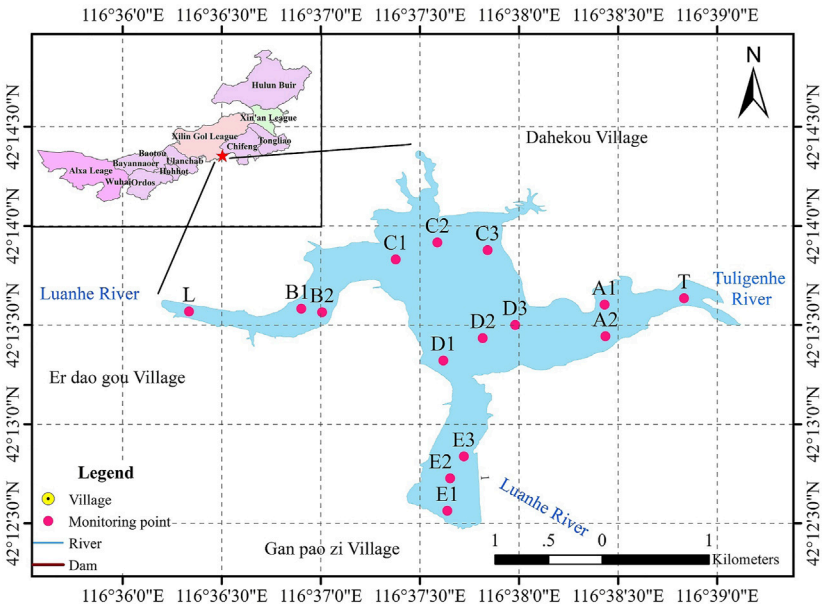


FIGURE 1 | Layout of the monitoring points in the Dahekou Reservoir.

the reservoir water has not been effectively controlled under these conditions although the external pollution sources are constantly being rectified and reduced. The eutrophication of the reservoir water has adversely affected the water quality of the lower reaches of the Luanhe River and has threatened the safety of the drinking water of the Tianjin residents in the lower reaches of the Luanhe River (Liu et al., 2010).

2.2 Sample Collection

Based on the Chinese water environment and the lake wetland survey norms, according to the distribution of the pollution sources, reservoir area, and water surface width of the Dahekou Reservoir, the reservoir was divided into five sections in the order of (A, B, C, etc.). A total of 18 water quality monitoring points were arranged (**Figure 1**).

From March 2018 to February 2019, each month, we traveled by hovercraft to the monitoring points using Global Positioning System (GPS) navigation. At each monitoring point, we measured the water depth (using a long-line portable LOWRANCE X-4 Pro water depth meter), the bottom water temperature, the dissolved oxygen content (using a YSI550A dissolved oxygen meter), the pH (using an HACH HQ30D instrument), and the flow rate (using a rotary cup flow meter). We measured the average value of the water temperature, the dissolved oxygen, and the pH at the bottom of each monitoring point as the basic physical and chemical index parameters for estimating the nitrogen release load of the sediments.

2.3 Subject Material

2.3.1 Sediment

The sediment samples collected from the bottom of the Dahekou Reservoir at a depth of 15 cm were packed in cloth bags, which were labeled and transported back to the laboratory in a cryogenic sealing device. The physical and chemical indexes of the sediment were determined to be as follows: a pH of 7.63, an organic matter content of 4.2%, and a total nitrogen (TN) content of 2246 mg/kg. The sediment samples were used as orthogonal simulation test samples to investigate the nitrogen release from the sediments.

2.3.2 Overlying Water

Samples of the water overlying the sediments at each sampling point were collected in polyethylene bottles and were transported back to the laboratory refrigerator (4°C) for storage. The TN concentration of the overlying water at the sampling point was 2.1 mg/L.

2.4 Test Scheme

2.4.1 Selection of the Factors Influencing the Orthogonal Test of the Total Nitrogen Exchange Flux in the Sediments

The sediment nitrogen release test method was conducted according to the "Lake Eutrophication Survey Specification." An orthogonal simulation test of the sediment-overlying water total nitrogen exchange was performed by selecting environmental factors, such as the temperature, pH, dissolved oxygen (DO), and hydrodynamics, which have significant effects on the sediment release.

1) Temperature: Na_2CO_3 and (1 + 5) H_2SO_4 solutions with 1 M concentrations were used to adjust the pH to 7.0. The TN concentration of the overlying water was measured in three triangular conical bottles placed in a biochemical incubator at 48-h interval under three ambient temperature conditions, that is, 5, 20, and 35°C.
2) pH: Na_2CO_3 and (1 + 5) H_2SO_4 solutions with 1 M concentrations were used to adjust the initial pH values of

TABLE 1 | Statistics of the total nitrogen released from the sediments of the Dahekou Reservoir in the orthogonal experiment.

Sequence number	Temperature (°C)	pH (dimensionless)	Dissolved oxygen (mg/L)	Hydrodynamic (r/min)	Release strength of two parallel tests (mg/m²•d)		Mean release intensity (mg/m²•d)	Range Ri
					1	2		
1	1 (5)	1 (6.0)	1 (2.0)	1 (0)	17.53	17.41	17.47	0.119
2	1 (5)	2 (7.5)	2 (4.0)	2 (30)	16.76	16.66	16.71	0.096
3	1 (5)	3 (9.0)	3 (6.0)	3 (60)	14.37	14.47	14.42	0.102
4	2 (20)	1 (6.0)	2 (4.0)	3 (60)	19.95	19.72	19.84	0.226
5	2 (20)	2 (7.5)	3 (6.0)	1 (0)	16.92	17.12	17.02	0.206
6	2 (20)	3 (9.0)	1 (2.0)	2 (30)	19.13	19.06	19.09	0.076
7	3 (35)	1 (6.0)	3 (6.0)	2 (30)	18.59	18.74	18.67	0.150
8	3 (35)	2 (7.5)	1 (2.0)	3 (60)	21.14	20.84	20.98	0.297
9	3 (35)	3 (9.0)	2 (4.0)	1 (0)	18.94	19.14	19.04	0.201
K_{j1}	48.60	54.71	57.54	53.53				
K_{j2}	55.95	52.55	55.59	54.47				
K_{j3}	58.70	55.98	50.11	55.24				

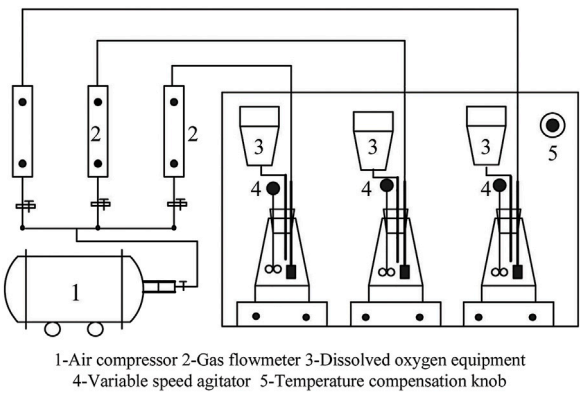


FIGURE 2 | Equipment used to test the nitrogen release from the sediments.

the overlying water samples to 6.0, 7.5, and 9.0. The TN concentrations of the overlying water samples were measured at 48-h interval in three simulated reactors.

3) DO: Na_2CO_3 and (1 + 5) H_2SO_4 solutions with 1 M concentrations were used to adjust the pH to 7.0. The DO concentration of the overlying water was maintained at 2.0 mg/L (N_2), 4.0 mg/L, and 6.0 mg/L by controlling the air or nitrogen aeration of the gas rotor flowmeter. The TN concentration of the overlying water was measured at 48-h interval in three reactors.
4) Hydrodynamic conditions: A biochemical incubator was used to control the ambient temperature at 25°C. Na_2CO_3 and (1 + 5) H_2SO_4 solutions with 1 M concentrations were used to adjust the pH of the overburden water of the reactor to 7.0. Three different hydrodynamic conditions (i.e., 0 r/min, 30 r/min, and 60 r/min) were set using an infinitely variable speed agitator. The TN concentration of the overlying water was measured at 48 h intervals in three reactors.

2.4.2 Orthogonal Experimental Design of the Nitrogen Release From Sediments

According to the orthogonal experimental design method, which has four factors and three levels [L_9 (3^4)], three groups of experimental designs, including nine schemes, were set up (**Table 1**). The orthogonal test device is shown in **Figure 2**.

2.5 Estimation of the Internal Nitrogen Release Pollution Load in the Sediments

Three groups of experiments were designed. Each group was run continuously for 14 days, and the sampling frequency was 1 time every 2 days. At the beginning of each experiment, 10 g of reservoir sediments were placed in glass bottles with plugs. Then, they were slowly injected into 500 ml of overlying water from the original reservoir, sealed with plugs, connected by pipelines, and placed in a biochemical constant temperature incubator.

Four factors including temperature, pH, DO, and water disturbance were selected as the influencing factors for the orthogonal simulation test of the total nitrogen exchange flux in the sediments. After the start of the experiment, every 2 days, 10 ml of the water overlying the sediment samples was taken using a pipette to determine the TN concentration. At the same time, the original calibration of the glass bottle was supplemented with 10 ml of surface water from the original reservoir. To ensure the rationality and reliability of the test data, two parallel tests were conducted for each test scheme (the error of the results of the two analyses was less than 5%), and the mean value of the results of the two tests was taken. The intensity of the total nitrogen release from the sediments was calculated using the following equation:

$$R = \left[V(C_n - C_0) + \sum_{i=1}^{n} V_i (C_{i-1} - C_a) \right] \Big/ A.$$

where R is the intensity of the nitrogen release (mg/m^2); V is the water sample's volume (L); C_n is the TN concentration of the water overlying the sediment for each sampling site (mg/L); C_0 is the initial TN concentration (mg/L); V_i is the volume of each sample (L); C_{i-1} is the TN concentration (mg/L); C_a is the TN concentration of the water added (mg/L); and A is the surface area of the sediment in contact with the water (m^2).

3 RESULTS AND DISCUSSION

3.1 Establishing and Testing the Model of the Intensity of the Nitrogen Release From the Sediments

3.1.1 Establishing a Mathematical Model of the Intensity of the Nitrogen Release From the Sediments in the Reservoir

According to the orthogonal simulation test results for the intensity of the total nitrogen release from the sediments, the release intensity from the sediments was calculated, and the statistics and range analysis were conducted (**Table 1**).

Based on the analysis of the data from the aforementioned nine orthogonal test schemes for investigating the intensity of the nitrogen release from the sediments, the relationship between the four environmental factors (i.e., temperature, pH, DO, and hydrodynamic conditions) and the intensity of the nitrogen release from the sediments was obtained using the multiple linear regression method:

$$R = 20.938 + 0.112[T] - 0.379[pH] - 0.621[DO] + 0.010[K],$$

where R is the intensity of the nitrogen release from the sediments ($mg/m^2 \bullet d$); T is the water temperature (°C), 5 < T < 35°C; pH is the pH of the water, 6 < pH < 9; DO is the dissolved oxygen content (mg/L), 0 < DO < 6 mg/L; and K is the speed of the stepless speed regulating agitator (r/min), 0 < K < 60.

3.1.2 Testing the Mathematical Model of the Intensity of the Nitrogen Release From the Sediments

The significance test for a multiple linear regression equation that is generally used to judge the overall significance of the regression model is the F test. If the statistical variable F is selected, then given the significance level a, the check F critical value is f1−2 = (k, n−k−1). If the statistical value F> f1−2 is obtained from the sample calculation, the regression model is considered to be significant. Otherwise, the regression model is not significant (Growns et al., 2003).

1) Significance test of the regression equation of the intensity of the nitrogen release from the sediments.

The significance of the regression equation of the mathematical model of the intensity of the nitrogen release from the sediments is tested: H_0:$\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$. The sum of the squared deviations of each factor is as follows:

$$S_T^2 = \sum y_i^2 - n\left(1/n \sum y_i\right)^2 = 61.302,$$

$$S_R^2 = \sum_{i=1}^{n}\left(\hat{y}_i - \bar{y}\right)^2 = 57.338,$$

$$S_E^2 = S_T^2 - S_R^2 = 61.302 - 57.338 = 3.964,$$

where S_T^2 is the sum of the squares of the total deviation; S_E^2 is the sum of the squares of the residual error; S_R^2 is the sum of the regression squares; and $n = 18$.Thus, $F = (S_R^2/4)/(S_E^2/(18-4-1)) = 47.016$,and $F = (S_R^2/4)/(S_E^2/(18-4-1)) = 47.016$.

For a significance level $\alpha = 0.05$, $F1-\alpha\,(4, 13) = 3.18$, $F > F1-\alpha$, and the rejection of H_0, the regression equation of the mathematical model of the intensity of the nitrogen release from the sediments was significant.

2) Significance test of the regression coefficient of the intensity of the nitrogen release from the sediments.

MATLAB 21.0 was used to test the significance of the regression coefficient of the model equation of the intensity of the nitrogen release from the sediments in the Dahekou Reservoir. The test results are presented in **Table 2**.

As can be seen from **Table 2**, the regression coefficients of the environmental factors T, pH, and DO in the regression equation of the mathematical model are highly significant, while the regression coefficients of the environmental factor K are not significant.

3.1.3 Comparison of the Significance Levels of the Factors Influencing the Intensity of the Nitrogen Release From the Sediments

A comparison of the significance levels of the various environmental factors affecting the nitrogen release from the sediments was carried out (**Figure 3**).

As is shown in **Figure 3**, as the temperature increased, the intensity of the nitrogen release from sediment increases obviously, and the sediments increased significantly. The increase in temperature accelerated the decomposition of the organic nitrogen in the sediments, and thus, the amount of organic nitrogen released into the overlying water increased. In addition, as the dissolved oxygen content decreased, the intensity of the nitrogen release from the sediments increased. This is because the surface layer and the water overlying the sediments are in an anoxic or anaerobic state when the dissolved oxygen content is low, so the release level of ammonia nitrogen, nitrate, and nitrite from sediments is significantly reduced, and the release intensity is much higher than 2–8 times the release rate under aerobic conditions (Zhao et al., 2018.). Ammonium nitrogen is the main form of internal nitrogen released, and its release degree is negatively correlated with the dissolved oxygen content, while it is positively correlated with temperature (Wang et al., 2011). The intensity of the nitrogen release from the sediments initially decreased and then increased as the pH changed from acidic to alkaline. This is attributed to the fact that when the pH of the solution was acidic, H^+ competed with NH_4^+ for adsorption positions on the colloidal particles of

TABLE 2 | Significant test results of the multiple linear regression coefficients of the TN release strength from the sediments.

Variance source deviation square sum degree of freedom variance F ratio significance	Deviation square sum	Degree of freedom	Variance	F ratio	Significance
T	33.9697	1	33.9697	111.4174	Highly significant
pH	3.8874	1	3.8874	12.7503	Highly significant
DO	18.5008	1	18.5008	60.681	Highly significant
K	0.98041	1	0.98041	3.2156	Not significant

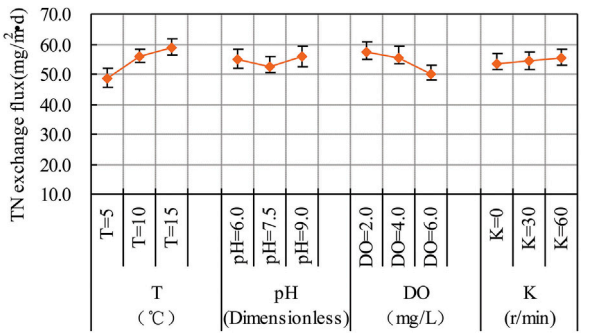


FIGURE 3 | Square deviation analysis of the total nitrogen released from the Dahekou sediments during the orthogonal experiments.

the sediments, leading to a significant increase in the intensity of the nitrogen release from the sediments. When the solution was close to neutral, the exchange capacity of NH_4^+ decreased correspondingly, and the intensity of the nitrogen release from the sediments was weakened. When the pH of the water was alkaline, NH_4^+ in the water reacted with OH^- to form molecular ammonia compounds, which escaped and increased the intensity of the nitrogen release from the sediments (Liu et al., 2014). As the hydrodynamic strength of the intensity of the nitrogen release from the sediments increased, the water disturbance not only resuspended the particle nitrogen in the sediments, but it also accelerated the nitrogen diffusion in interstitial water in the sediments, which increased the material exchange between the overlying water and the sediments, effectively promoting the release of nitrogen from the sediments.

To compare the order of the influences of the various environmental factors on the intensity of the nitrogen release from the sediments, range analysis was conducted on the orthogonal test data for the nitrogen release from the sediments, and the results are as follows:

Temperature: $S_{jA}^2 = 1/3 \times (K_{j1}^2 + K_{j2}^2 + K_{j3}^2) - \frac{1}{3\times 3}(a_1 + a_2 + ...a_9) = 18.54$.

In the formula, K_{j1}, K_{j2}, and K_{j3} are the sums of test values under various levels; a_1, a_2.....a_9 is the value of the total nitrogen release intensity; and S_{jA}, S_{jB}, S_{jC}, and S_{jD} are the standard variances under the environmental conditions of temperature, pH, DO, and K, respectively.

In the same way: $S_{jB}^2 = 2.00$, $S_{jC}^2 = 8.98$, and $S_{jD}^2 = 0.49$.

The results of the range analysis of the data show that within the range of values specified for the various environmental factors, there is a relationship of $S_{jA}{}^2 > S_{jC}{}^2 > S_{jB}{}^2 > S_{jD}{}^2$. In other words, the significance of the four environmental factors on the release of nitrogen from the sediments in the Dahekou Reservoir is ranked as follows: water temperature (T) > dissolved oxygen (DO) > pH value > hydrodynamic (K).

3.2 Estimation and Variation Analysis of the Internal Nitrogen Release Pollution Load in the Sediments

According to the monitoring data for the actual physical and chemical indexes, such as water temperature, pH, and dissolved oxygen, at the bottom of the Dahekou Reservoir in 2018, the intensity of the nitrogen release from the reservoir sediments during different time periods was estimated using the mathematical equation:

$$W = \sum R_i \cdot \Delta T_i \cdot A \times 10^{-9},$$

where W is the internal nitrogen release pollution load of the reservoir sediments (t); R_i is the nitrogen exchange flux under ice conditions [mg/(m^2•d)]; ΔT_i is the corresponding time period under ice conditions (d); and A is the corresponding reservoir area at different time periods (m^2).

The Dahekou Reservoir is located in a windy and sandy area, with an average annual wind speed of >3.3 m/s. The kinetic energy generated by the wind-wave disturbances and the scouring effect of the discharged water on the sediments during power generation is transferred to the sediment-water interface, from the sediment-water interface to the sediments, and the sediments are resuspended, which has an important effect on the migration and transformation of phosphorus nutrient salt between the sediment and the water (Schelske et al., 1974). The wave hydraulic effect of the reservoir sediment disturbance intensity is currently not directly measurable using the currently available experimental instruments. Therefore, in this study, a Nanjing south water LB70-1 c cup type current meter was used to determine the value for the reservoir. Based on the relationship between the flow velocity and the spin cup speed, the hydraulic disturbance intensity was indirectly determined using the established mathematical model of the sediment

TABLE 3 | Estimation of the TN release pollution load from the sediments in the Dahekou Reservoir.

Month	Time (d)	Water temperature (°C)	pH (dimensionless)	DO (mg/L)	Hydrodynamic force (r/min)	Exchange flux ($mg/m^2 \bullet d$)	Reservoir area (km^2)	TN release (t/month)
January	31	5.48	7.81	8.03	12.40	13.73	2.31	0.983
February	29	5.66	7.70	7.72	16.81	14.03	2.53	1.029
March	31	6.85	8.08	6.69	22.45	14.71	2.47	1.127
April	30	8.38	8.09	6.72	24.34	14.88	2.54	1.134
May	31	9.58	8.29	5.45	32.45	15.81	2.67	1.309
June	30	13.09	6.18	4.65	21.45	17.39	2.71	1.414
July	31	14.46	6.26	2.34	20.48	18.94	2.14	1.256
August	31	15.66	6.49	2.08	16.36	19.10	2.40	1.421
September	30	11.72	6.97	5.53	23.34	16.41	2.87	1.413
October	31	9.53	6.68	6.49	18.65	15.63	2.34	1.134
November	30	8.97	7.34	7.32	16.75	14.78	2.30	1.020
December	31	5.68	7.62	8.11	13.82	13.79	2.43	1.039
Total								14.278

TABLE 4 | Estimations of the reservoir's TN pollution load from runoff.

Month	Time(d)	Luanhe River			Tuligen River			Total TN load (t/month)
		Volume of runoff (m^3/s)	Storage TN concentration (mg/L)	TN load (t/month)	Volume of runoff) m^3/s)	Storage TN concentration (mg/L)	TN load (t/month)	
January	31	0.061	3.54	0.578	0.031	2.81	0.233	0.812
February	28	0.06	3.37	0.489	0.028	2.66	0.180	0.669
March	31	0.071	3.55	0.675	0.033	2.58	0.228	0.903
April	30	0.103	4.01	1.071	0.039	4.21	0.426	1.496
May	31	0.128	9.01	3.089	0.042	8.45	0.951	4.040
June	30	0.121	5.13	1.609	0.044	5.32	0.607	2.216
July	31	0.143	10.33	3.957	0.056	8.46	1.269	5.225
August	31	0.167	8.53	3.815	0.048	7.53	0.968	4.783
September	30	0.192	9.12	4.539	0.061	8.95	1.415	5.954
October	31	0.199	11.32	6.034	0.057	10.43	1.592	7.626
November	30	0.112	5.17	1.501	0.036	3.6	0.336	1.837
December	31	0.072	2.57	0.496	0.029	2.41	0.187	0.683
Total				27.852			8.392	36.244

nitrogen release for the pollution load estimation. The results are presented in **Table 3**.

$$v = a + bn,$$

where v is the flow rate (m/s); a is the current meter constant (0.0067 m) (m/s); b is the hydraulic pitch of the rotary cup (0.7432 m) (m); and n is the rotor velocity of the current meter (s^{-1}).

As can be seen from **Tables 3**, **4**, the amount of TN released from the sediments in the Dahekou Reservoir was 14.278 t/a in 2018, among which the amount of TN released during the wet season (from June to September) was 5.504 t, accounting for 38.55% of the annual amount of TN released. From February to May, the total nitrogen released from the sediments was 4.599 t, accounting for 32.21% of the total nitrogen released throughout the year. During the dry season (October to January), the total nitrogen released from the sediments was 4.176 t, accounting for 29.25% of the total nitrogen released throughout the year. The annual amount of total nitrogen released from the sediments in the reservoir was the largest in the wet season and the smallest in the dry season, which is consistent with the results reported by Liu et al. (2006). Based on the comprehensive analysis of the water temperature, pH, DO, and the hydrodynamic conditions, the various physical, chemical, and biological reaction conditions (e.g., diffusion and organic matter degradation) during the wet season are conducive to promoting and accelerating the decomposition of the organic nitrogen in the sediments (Wang et al., 2013; Zhu et al., 2017), which leads to an increase in the amount released into the overlying water. The level period after the reservoir freezes is influenced by reservoir fisheries and power stations and by human activities, such as water disturbances. This can cause sediment resuspension and the particulate nitrogen in the accelerated interstitial diffusion of nitrogen from the sediment into the water, which increases the material exchange capacity between overlying water and sediment, effectively promoting the release of nitrogen from the sediments (Nowlin et al., 2005). However, the level period also includes algae growth, photosynthesis, and increase in pH, which is beneficial to sediment nitrogen adsorption by the upper water, so the plentiful release of nitrogen from the sediment is not

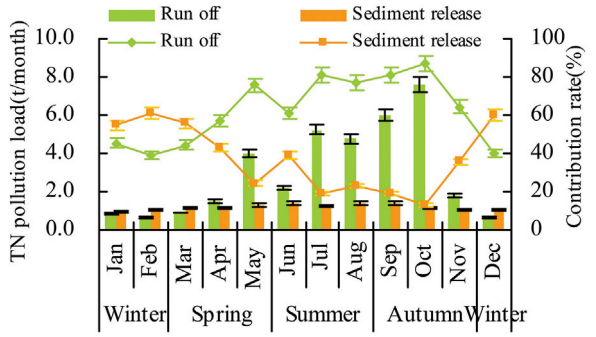


FIGURE 4 | Monthly variations in the reservoir's TN contribution rate from two sources.

very obvious (Wu and Hua, 2014). During the dry, deep freeze period, the water temperature is much lower than at other times throughout the year. The use of NO_2^--N by water algae microorganisms or dormant microorganisms (Robarts and Zohary, 1987; Cao et al., 2008) and the mud microbial activities near the weak interface is slow, and thus, the mud nutrient exchange flux is tiny at the interface, and the nitrogen release from the sediments is not significant (Nowlin et al., 2005; Cheng et al., 2015). As a result, the nitrogen release intensity reaches a minimum for the year during this period, and for the sediments in the reservoir, the contribution rate of the nitrogen nutrient released is smaller.

3.3 Identification of the Contribution of Nitrogen Pollution From the Sediments in the Reservoir

To identify the contribution rate of the internal nitrogen input to reservoir eutrophication, based on the 2018 storage in the Luanhe River mouth section, the water quality monitoring data for the spit storage root river mouth section, the Duolun county big mouth spit force provided by the river, and the data for the incoming runoff at the Luanhe River reservoir management stations, the same period for the contribution of the rivers to the TN pollution load was estimated using the following equation (**Table 4**):

$$M = \rho_i \times Q_i \times 3600 \times 24 \times n_i \times 10^{-6},$$

where M is the monthly TN pollution load of the river runoff input into the storage area (t/month); Q_i is the monthly river inflow cross-section flow (m^3/s); and N_i is the number of days per month (d).

As can be seen from **Tables 3**, **4**, the Dahekou Reservoir's TN input pollution load was 36.244 t in 2018, and the internal nitrogen released from the sediment into the nitrogen load was 14.278 t, accounting for 28.26% of the reservoir TN pollution load. In the similarly eutrophic Lake Taihu, China, based on the relationship between the sediment resuspension flux and the wind speed, Pang et al. (2008) estimated the amount of internal nitrogen released from the Taihu lake sediments and determined that it accounted for 25.7% of the total exogenous nitrogen input.

To determine the contribution rate of the sediment to the TN pollution of the Dahekou Reservoir, the calculation results for the sediment pollution load and the TN pollution load of the river runoff into the reservoir were plotted (**Figure 4**) to compare their contributions.

As is shown in **Figure 4**, the contributions of the nitrogen released from the sediments and that input by the river runoff to the total nitrogen pollution load exhibit regular trends. In spring (March–May), when the reservoir is unblocked, the water temperature is affected by stratification and inversion, and the nitrogen nutrients in the sediment have significant release characteristics (Wang et al., 2019), which leads to the significant release, adsorption, and desorption of the nitrogen and phosphorus in the sediments, including suspended solids (Gao et al., 2013; Testa et al., 2013). Therefore, the nutritional level of the overlying water body is increased, and it is affected by biological activities, such as algae and zooplankton growth and the double impact of the hydropower generation discharge of the Datang International Duolun Hydropower Station on the disturbance of the reservoir sediment. In summer (June–August), the temperature and rainfall intensity increase sharply, and the reservoir river runoff also increases soaring. Because the reservoir is surrounded by fertilized farmland and the rainfall surface runoff from this farmland flows into the Luanhe River and thus into the reservoir, the contribution rate of the total nitrogen pollution load increased to 61.05–80.62%, the river runoff nitrogen nutrient transport has become the main input to the reservoir in the summer, and the total nitrogen content increased. In autumn (September–November), because of the increase in rainfall in the areas upstream of the reservoir, especially in October, the river input increased rapidly, the inflow of surface runoff reached a peak for the year, the water level rose significantly, and the rivers' total nitrogen pollutant load input to the TN storage reached the maximum value of 7.626 t/month, accounting for 87.05% of the monthly total nitrogen storage load. For the nitrogen released from the sediments in the reservoir, the contribution rate of the total nitrogen reached the annual minimum, and the contribution of the river runoff nitrogen nutrient transport to the reservoir's water quality and eutrophication level played a key role. In winter (December–February), the reservoir entered the ice sealing period, the runoff from the Luanhe and the Tuligen rivers into the reservoir system reached the lowest level for the year. Furthermore, the point source and non-point source pollution along the rivers were reduced, and the water quality improved significantly, resulting in the reduction of the contribution rate of the total nitrogen pollution into the reservoir. Under the influence of the lower water temperature and higher dissolved oxygen content, the activity of microorganisms also decreased further, and the rate of nitrification and denitrification reached the lowest of the year; thus the intensity of the nitrogen release from the sediments reached the lowest for the year. Because the ice sealing period occurred in the dry season, the area of the reservoir was reduced, and the total nitrogen pollution load released from the sediments was lower than the total nitrogen pollution load at other times of the year. However, owing to the huge decrease in the river runoff in the dry season, the

contribution rate of the total nitrogen released from the sediments to the total reservoir pollution load was higher than in other seasons, reaching the highest level (54.76–60.60%), which is higher than the total nitrogen pollution load the rivers contribute to the reservoir. The total nitrogen released from the sediments becomes an important source of nitrogen nutrition when the reservoir is frozen.

4 CONCLUSION

Nitrogen release from the sediments in a reservoir (lake) is a very complex dynamic process, which is influenced by multiple factors. To determine the factors influencing the nitrogen release from the sediment, the intensity of the internal nitrogen release and its contribution to the pollution of the reservoir were estimated. An orthogonal test of nitrogen release from sediments for four factors and three factors [$L_9(3^4)$] were designed, laboratory simulation tests of the influences of the four factors (i.e., T, pH, DO, and K) were conducted, and the test results were analyzed. The following conclusions were drawn.

1) The mathematical model equation [R = 20.938 + 0.122 (T)−0.379 (pH)−0.621 (DO) + 0.010 (K)] for the nitrogen release from the sediments for the ranges of the specified environmental variables was established based on the results of laboratory orthogonal simulation experiments, and the significance of the model was determined using the F test. The results revealed that the regression equation of the mathematical model of the intensity of the nitrogen release from the sediments is significant, and the regression coefficients of the environmental factors T, pH, and DO were highly significant, but that of K was not.
2) In 2018, the TN release load of the sediments in the Dahekou reservoir was 14.278 t/a, of which the total nitrogen released from the sediments in the wet season (from June to September) was 5.504 t, accounting for 38.55% of the total nitrogen released throughout the year. The total nitrogen released from the sediments in the horizontal period (from February to may) was 4.599 t, accounting for 32.21% of the total nitrogen releases throughout the year. During the dry season (October to January), the total nitrogen released from the sediment was 4.176 t, accounting for 29.25% of the total nitrogen released throughout the year.
3) In spring, the contribution rate of the nitrogen released from the sediments and that of the river runoff to the nitrogen and nutrients in the reservoir were relatively close, so they were the main sources of nutrients for the reservoir water during this season. In summer and autumn, although the environmental conditions were conducive to the release of nitrogen nutrients from the sediments into the overlying water, the contribution rate of the nitrogen released from the sediments into the overlying water decreased to 19.38–38.95% due to the increased contribution of the runoff pollution from the rivers upstream. The contribution rate of the nitrogen released from the sediments reached the lowest level at this time, and the input of nitrogen nutrients by the upstream river was the main cause of the eutrophication of the reservoir. During the frozen period, the contribution rate of the nitrogen released from the sediments into the overlying water reached the highest level (54.76–60.60%), and it was one of the main sources of nitrogen nutrients in the water.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material; further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

JL conceived of the presented idea and developed the theory and performed the computations. XS verified the analytical methods. TL supervised the findings of this work. All authors discussed the results and contributed to the final manuscript.

FUNDING

This study was funded by the Key Laboratory of Water Resource Protection and Utilization in the Inner Mongolia Autonomous Region, the National Natural Science Foundation of China (51669023 and 51869021), and the National Key Research and Development Program of China (2019YFC0609204).

REFERENCES

Beeton, A. M. (1965). Eutrophication of the St. Lawrence Great Lakes1. *Limnol. Oceanogr.* 10 (2), 240–254. doi:10.4319/lo.1965.10.2.0240

Bocrs, P. C. M., and Istvanovics, V. (1991). The Influence of pH on Phosphate Release from Lake Sediments. *Water Res.* 25 (3), 309–311. doi:10.1016/0043-1354(91)90010-N

Cao, H.-S., Tao, Y., Kong, F.-X., and Yang, Z. (2008). Relationship between Temperature and Cyanobacterial Recruitment from Sediments in Laboratory and Field Studies. *J. Freshw. Ecol.* 23 (3), 405–412. doi:10.1080/02705060.2008.9664217

Cheng, X., Hou, L., Liu, M., Zheng, Y., Yin, G., Li, X., et al. (2015). Inorganic Nitrogen Exchange across the Sediment-Water Interface in the Eastern Chongming Tidal Flat of the Yangtze Estuary. *Environ. Earth Sci.* 74 (3), 2173–2184. doi:10.1007/s12665-015-4207-z

Conley, D. J., Humborg, C., Rahm, L., Savchuk, O. P., and Wulff, F. (2002). Hypoxia in the Baltic Sea and Basin-Scale Changes in Phosphorus Biogeochemistry. *Environ. Sci. Technol.* 36 (24), 5315–5320. doi:10.1021/es025763w

Davis, C. C. (1964). Evidence for the Eutrophication of Lake Erie from Phytoplankton Records. *Limnol. Oceanogr.* 9 (3), 275–283. doi:10.4319/lo.1964.9.3.0275

Gao, L. I., Zhang, L., Hou, J., Wei, Q., Fu, F., Shao, H., et al. (2013). Decomposition of Macroalgal Blooms Influences Phosphorus Release from the Sediments and Implications for Coastal Restoration in Swan Lake, Shandong, China. *Ecol. Eng.* 60 (6), 19–28. doi:10.1016/j.ecoleng.2013.07.055

Gelin, C. (1975). Nutrients, Biomass and Primary Productivity of Nannoplankton in Eutrophic Lake Vombsjön, Sweden. *Oikos* 26, 121–139. doi:10.2307/3543701

Growns, I., Gehrke, P. C., Astles, K. L., and Pollard, D. A. (2003). A Comparison of Fish Assemblages Associated with Different Riparian Vegetation Types in the Hawkesbury-Nepean River System. *Fish. Manag. Ecol.* 10 (4), 209–220. doi:10.1046/j.1365-2400.2003.00337.x

Gu, X.-z., Chen, K.-n., Zhang, L., and Fan, C.-x. (2016). Preliminary Evidence of Nutrients Release from Sediment in Response to Oxygen across Benthic Oxidation Layer by a Long-Term Field Trial. *Environ. Pollut.* 219, 656–662. doi:10.1016/j.envpol.2016.06.044

Hammer, U., Walker, K., and Williams, W. (1973). Derivation of Daily Phytoplankton Production Estimates from Short-Term Experiments in Some Shallow, Eutrophic Australian Saline Lakes. *Mar. Freshw. Res.* 24 (3), 259–266. doi:10.1071/mf9730259

Han, H., Lu, X., Burger, D. F., Joshi, U. M., and Zhang, L. (2014). Nitrogen Dynamics at the Sediment-Water Interface in a Tropical Reservoir. *Ecol. Eng.* 73, 146–153. doi:10.1016/j.ecoleng.2014.09.016

Hartmann, J., and Nümann, W. (1977). Percids of Lake Constance, a Lake Undergoing Eutrophication. *J. Fish. Res. Bd. Can.* 34 (10), 1670–1677. doi:10.1139/f77-231

Holdren, G. C., and Armstrong, D. E. (1980). Factors Affecting Phosphorus Release from Intact Lake Sediment Cores. *Environ. Sci. Technol.* 14 (1), 79–87. doi:10.1021/es60161a014

Liu, Y., Zhang, Z., and Duan, X. J. (2006). Prediction Model and the Control of Nitrogen Releasing from Sediment of Lakes, China. *J. Agro-environment Sci.* 25 (6), 1603–1606.

Liu, J., Chen, Q., and Li, Y. (2010). Ecological Risk Assessment of Water Environment for Luanhe River Basin Based on Relative Risk Model. *Ecotoxicology* 19 (8), 1400–1415. doi:10.1007/s10646-010-0525-9

Liu, X., Xian, P., Wen, Z. W., Xiang, T., Yang, L. H., and Guo, M. F. (2014). "Research on the Influence of Overlying Water pH to the Sediment Release," in *Environmental Technology and Resource Utilization Ii*. Editor L. Zhangs (Trans Tech Publications Ltd Stafa-Zurich), 556–562.

Mayer, T., Rosa, F., and Charlton, M. (2005). Effect of Sediment Geochemistry on the Nutrient Release Rates in Cootes Paradise Marsh, Ontario, Canada. *Aquatic Ecosyst. Health & Manag.* 8 (2), 133–145. doi:10.1080/14634980590954986

Ni, Z., and Wang, S. (2015). Historical Accumulation and Environmental Risk of Nitrogen and Phosphorus in Sediments of Erhai Lake, Southwest China. *Ecol. Eng.* 79, 42–53. doi:10.1016/j.ecoleng.2015.03.005

Nowlin, W. H., Evarts, J. L., and Vanni, M. J. (2005). Release Rates and Potential Fates of Nitrogen and Phosphorus from Sediments in a Eutrophic Reservoir. *Fresh Water Biol.* 50 (2), 301–322. doi:10.1111/j.1365-2427.2004.01316.x

Okino, T., and Kato, K. (1987). Lake Suwa-Eutrophication and its Partial Recent Recovery. *Geo J.* 14 (3), 373–375. doi:10.1007/bf00208212

Pang, Y., Yan, R. R., Yu, Z. B., Li, Y. P., and Li, R. L. (2008). Suspension-sedimentation of Sediment and Release Amount of Internal Load in Lake Taihu Affected by Wind. *Huan Jing Ke Xue* 29 (9), 2456–2464.

Pennuto, C. M., Burlakova, L. E., Karatayev, A. Y., Kramer, J., Fischer, A., and Mayer, C. (2014). Spatiotemporal Characteristics of Nitrogen and Phosphorus in the Benthos of Nearshore Lake Erie. *J. Gt. Lakes. Res.* 40 (3), 541–549. doi:10.1016/j.jglr.2014.02.013

Robarts, R. D., and Zohary, T. (1987). Temperature Effects on Photosynthetic Capacity, Respiration, and Growth Rates of Bloom-forming Cyanobacteria. *N. Z. J. Mar. Freshw. Res.* 21 (3), 391–399. doi:10.1080/00288330.1987.9516235

Schelske, C. L., Rothman, E. D., Stoermer, E. F., and Santiago, M. A. (1974). Responses of Phosphorus Limited Lake Michigan Phytoplankton to Factorial Enrichments with Nitrogen and Phosphorus1. *Limnol. Oceanogr.* 19 (3), 409–419. doi:10.4319/lo.1974.19.3.0409

Schindler, D. W. (1974). Eutrophication and Recovery in Experimental Lakes: Implications for Lake Management. *Science* 184 (4139), 897–899. doi:10.1126/science.184.4139.897

Shan, L., He, Y., Chen, J., Huang, Q., Lian, X., Wang, H., et al. (2015). Nitrogen Surface Runoff Losses from a Chinese Cabbage Field under Different Nitrogen Treatments in the Taihu Lake Basin, China. *Agric. Water Manag.* 159, 255–263. doi:10.1016/j.agwat.2015.06.008

Testa, J. M., Brady, D. C., Di Toro, D. M., Boynton, W. R., Cornwell, J. C., and Kemp, W. M. (2013). Sediment Flux Modeling: Simulating Nitrogen, Phosphorus, and Silica Cycles. *Estuar. Coast. Shelf Sci.* 131, 245–263. doi:10.1016/j.ecss.2013.06.014

Wang, Z., Zhao, L., Li, X., Ma, K., Huang, X., Feng, Z., et al. (2011). The Factors Affecting the Intensity of Lake Endogenous Nitrogen Release under the Conditions of Different Time Scales, China. *J. Agro-Environment Sci.* 30 (12), 2542–2547.

Wang, P., Shen, X., Wang, S., and Hao, S. (2013). Remobilization of Phosphorus from Sediments of Taihu Lake during Periods of Simulated Resuspension. *Water Environ. Res.* 85 (11), 2209–2215. doi:10.2175/106143013x13736496909590

Wang, Y. M., Li, K. F., Liang, R., Han, S., and Li, Y. (2019). Distribution and Release Characteristics of Phosphorus in a Reservoir in Southwest China. *Int. J. Environ. Res. Public Health* 16 (3), 12. doi:10.3390/ijerph16030303

Wu, D., and Hua, Z. (2014). The Effect of Vegetation on Sediment Resuspension and Phosphorus Release under Hydrodynamic Disturbance in Shallow Lakes. *Ecol. Eng.* 69, 55–62. doi:10.1016/j.ecoleng.2014.03.059

Xiang, S., Nie, F., Wu, D., and Liu, X. (2015). Nitrogen Distribution and Diffusive Fluxes in Sediment Interstitial Water of Poyang Lake. *Environ. Earth Sci.* 74 (3), 2609–2615. doi:10.1007/s12665-015-4281-2

Xue, W., and Lu, S-Y. (2015). Effects of Inactivation Agents and Temperature on Phosphorus Release from Sediment in Dianchi Lake, China. *Environ. Earth Sci.* 74 (5), 3857–3865. doi:10.1007/s12665-014-3910-5

Yan, R. R., Pang, Y., Chen, X. F., Zhao, W., and Ma, J. (2008). Effect of Disturbance on Growth of Microcystis Aeruginosa in Different Nutrient Levels. *Huan Jing Ke Xue* 29 (10), 2749–2753.

Yang, W.-Q., Xiao, H., Li, Y., and Miao, D.-R. (2017). Vertical Distribution and Release Characteristics of Nitrogen Fractions in Sediments in the Estuaries of Dianchi Lake, China. *Chem. Speciat. Bioavailab.* 29 (1), 110–119. doi:10.1080/09542299.2017.1352460

Yang, J.-F., Xu, P., Wu, S.-L., Liu, Z.-P., He, M., Zheng, L.-Y., et al. (2018). Distribution, Sources and Ecological Risk Assessment of Heavy Metals in the Surface Sediments of Lake Liuye and its Adjacent Waters, China. *J. Radioanal. Nucl. Chem.* 318 (2), 1131–1142. doi:10.1007/s10967-018-6091-y

Zhang, K., Cheng, P.-d., Zhong, B.-c., and Wang, D.-z. (2012). Total Phosphorus Release from Bottom Sediments in Flowing Water. *J. Hydrodyn.* 24 (4), 589–594. doi:10.1016/s1001-6058(11)60281-3

Zhao, H., Zhang, L., Wang, S., and Jiao, L. (2018). Features and Influencing Factors of Nitrogen and Phosphorus Diffusive Fluxes at the Sediment-Water Interface of Erhai Lake. *Environ. Sci. Pollut. Res.* 25 (2), 1933–1942. doi:10.1007/s11356-017-0556-3

Zhong, J.-C., You, B.-S., Fan, C.-X., Li, B., Zhang, L., and Ding, S.-M. (2008). Influence of Sediment Dredging on Chemical Forms and Release of Phosphorus. *Pedosphere* 18 (1), 34–44. doi:10.1016/s1002-0160(07)60100-3

Zhu, H. W., Cheng, P. D., Li, W., Chen, J. H., Pang, Y., and Wang, D. Z. (2017). Empirical Model for Estimating Vertical Concentration Profiles of Re-suspended, Sediment-Associated Contaminants. *Acta Mech. Sin.* 33 (5), 846–854. doi:10.1007/s10409-017-0650-2

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Environmental Science

TYPE Original Research
PUBLISHED 16 August 2022
DOI 10.3389/fenvs.2022.831122

OPEN ACCESS

EDITED BY
Chuanfu Zang,
South China Normal University, China

REVIEWED BY
Weili Duan,
Xinjiang Institute of Ecology and Geography (CAS), China
Dengfeng Liu,
Xi'an University of Technology, China

*CORRESPONDENCE
Yi He,
yihe@nwu.edu.cn
Xingmin Mu,
963916337@qq.com

SPECIALTY SECTION
This article was submitted to Environmental Informatics and Remote Sensing,
a section of the journal
Frontiers in Environmental Science

RECEIVED 08 December 2021
ACCEPTED 11 July 2022
PUBLISHED 16 August 2022

CITATION
Zhang Y, He Y, Mu X, Jia L and Li Y (2022), Base-flow segmentation and character analysis of the Huangfuchuan Basin in the middle reaches of the Yellow River, China.
Front. Environ. Sci. 10:831122.
doi: 10.3389/fenvs.2022.831122

Base-flow segmentation and character analysis of the Huangfuchuan Basin in the middle reaches of the Yellow River, China

Yaru Zhang[1,2], Yi He[1,2,3]*, Xingmin Mu[3,4]*, Liping Jia[1,2] and Yanlin Li[1,2]

[1]Shaanxi Key Laboratory of Earth Surface System and Environmental Carrying Capacity, College of Urban and Environmental Sciences, Northwest University, Xi'an, Shaanxi, China, [2]Yellow River Institutes of Shaanxi Province, Xi'an, Shaanxi, China, [3]The Research Center of Soil and Water Conservation and Ecological Environment, Chinese Academy of Sciences and Ministry of Education, Yangling, Shaanxi, China, [4]Institute of Soil and Water Conservation, Northwest A&F University, Yangling, China

In recent years, with the deterioration of the ecological environment, runoff in the Yellow River has been decreasing. Given these phenomena, it is necessary to research the base-flow segmentation methods and its characteristics. The Huangfuchuan Basin in the middle reaches of the Yellow River basin was selected as the research area. This paper calculated the base-flow with commonly used base-flow segmentation methods, including Base-flow index method (BFI), Hydrograph-separation techniques (HYSEP), and the digital filtering methods, and compared the applicability of these methods in the study area. Then the variation characteristics, abrupt change year, periodic change, and future trend of the base-flow were analyzed. The results are as follows: 1) Through the analysis and comparison of several common base-flow segmentation methods, these methods had a marked difference in base-flow segmentation. The variance and extreme-ratio of DF4 in the digital filtering methods were small. The correlation between DF4 and other methods was high. DF4 in the digital filtering methods was the most appropriate method in the research area. 2)The runoff and base-flow index presented a decreasing trend. The annual average runoff was 1,100.27 × 10^4 m^3, and the maximum base-flow was 0.21. The base-flow also showed a significant decrease trend by the Mann-Kendall trend test, the average base-flow was 10,578.35 × 10^4 m^3. 3) The base-flow had periodic variations of 3–6, 7–18, and 19–32. In the 19–32 years time scale, three oscillations alternated between abundant and dry. The whole time was 31 years as the center of the periodic change, the first primary cycle of base-flow change. 4) The base-flow mutation occurred in 1986. The Hurst index of base-flow in the Huangfuchuan Basin was 0.84. It was shown that the trend of future development is positively correlated with past change, showing a trend of continuous decrease. Selecting the appropriate method of base-flow segmentation and reasonably analyzing the variation characteristics of base-flow can provide scientific guidance for the ecological environment construction and water resource evaluation in the Huangfuchuan Basin and even the Yellow River Basin.

KEYWORDS

base-flow, the Huangfuchuan basin, base-flow segmentation, wavelet analysis, Mann-Kendall test

Introduction

Water resources are the primary condition for the sustainable development of the Earth's ecological environment. Meanwhile, it is an indispensable material for the survival of organisms (Liao et al., 2020). With the rapid growth of the social economy, many countries face different degrees of water shortages (Chang, 2009). Climate change and human activities pose significant threats to water resources and water cycles. The complex relationship between the water cycle, climate change, and human activities has become the focus of scientific research today (Bastin et al., 2019).

Base-flow is a considerable part of water resources and plays a vital role in the water cycle (Chen et al., 2006). Base-flow is the essential runoff replenished by groundwater, and it is also the primary runoff to maintain the stability of the ecosystem (Brutsaert and Nieber, 1977; Tallaksen, 1995). It has the characteristics of stable flow and slight variation in annual distribution. The distribution features of base-flow are influenced by climate type, topography, and soil properties. Meanwhile, the influence of human activities is also gradually growing (Veldkamp et al., 2017). Base-flow plays a crucial role in maintaining the water balance, managing water resources reasonably and scientifically, and maintaining ecological river health. Base-flow is a hot topic that attracts attention and exploration in hydrology research (Lin and Li, 2010). In recent years, experts and scholars have performed much research on the base-flow and obtained some breakthroughs, but the definition of base-flow has not been unified, which has led to the diversity of base-flow segmentation methods (Santhi et al., 2008; Janke et al., 2014).

Base-flow segmentation is the focus and difficulty of current base-flow research. To date, many base-flow segmentation methods have been proposed by scholars from various countries, but it is still difficult to find a universally recognized base-flow segmentation method (Xu et al., 2016; Eckhardt, 2008). The traditional techniques of base-flow segmentation include the straight-line segmentation method and regression curve method. Nevertheless, these methods are difficult to apply to long time series data due to their complex calculation (Zuo et al., 2007). With the wide application of computer technology in hydrology, automatic base-flow segmentation methods have been developed, including BFI (Wels et al., 1991), HYSEP (Sloto and Crouse, 1996), digital filter methods (Arnold et al., 1995), Kalinin method (Wittenberg and Sivapalan, 1999), hydrological model method, and isotope method (Li et al., 2009). The United States Geological Survey (USGS) developed the Base-flow Partitioning Method (PART) and compared it with the manual mode. The difference between the result of this method and manual calculation was less than 10%, so the result was reasonable (Rutledge, 1993). The digital filtering method was first proposed in 1979. Nathan and McMahon improved the technique and verified the base-flow segmentation results of different watersheds. The segmentation results of the digital filtering method were reasonable. Automatic base-flow segmentation technology has become mature and widely used (Lyne and Hollick, 1979; Chapman, 1991). The smoothed minimum method proposed by the British Institute of Hydrology was compared with the manual method to segment the base-flow. A better result can be obtained when the optimal inflection point test factor is 0.9 (Mazvimavi et al., 2004). With the wide application of 3S technology, the emergence of distributed hydrological models, such as the TOPgraphy based hydrotogical Model (TOPMODEL) and Soil and Water Assessment Tool (SWAT), provided a suitable method for base-flow segmentation (Partington et al., 2012). These methods had reasonable objectivity and repeatability; consequently, they had been applied and developed rapidly in current studies. However, the segmentation results of these methods were not identical or even entirely different.

In practical applications, selecting stable and reliable base-flow segmentation according to hydrogeological conditions of different basins has become the research focus. Eckhardt used seven other methods to calculate the base-flow of 65 watersheds in the northern United States. The results showed that the calculation results based on the two-parameter Eckhardt filtering method were more reasonable and more consistent with the receding process of base-flow (Eckhardt, 2008). Zhang et al. used the SWAT model to simulate surface runoff and base-flow in a small watershed in the United States. The results showed that SWAT model simulation had superior performance. This technology can be used as a valuable tool to explore surface runoff and base-flow. It also provided a reference for further diagnostic evaluation and model recognition (Zhang et al., 2011). Wang et al. used five automatic base-flow segmentation methods to segment the runoff data of Luoshan Station in the Yangtze River from 1965 to 2012. They found that the annual base-flow process segmented by the five methods was significantly different. The calculation results of BFI were more in line with the various characteristics of base-flow. It was a suitable method for base-flow segmentation in the middle reaches of the Yangtze River (Wang et al., 2015). Lv et al. used the fixed interval method, sliding time method, local minimum method, and digital filtering method to calculate base-flow in the Fen River Basin. The curve of the digital filtering method was smooth and consistent with runoff, which conformed to the regularity of base-flow. The digital filtering method can objectively reflect the base-flow condition of the Fenhe River basin (Lv et al., 2021). Yang *et al.* used the SWAT model to simulate the base-flow of the Luohe River Basin. The results showed that the accuracy of the SWAT model simulation is higher than that of the digital filtering methods (Yang et al., 2003).

The Yellow River is an essential river in northern China. The Huangfuchuan Basin is one of the most severe areas of soil erosion in the middle of the Yellow River Basin. Climate change, plant change, and soil and water conservation measures have significantly affected

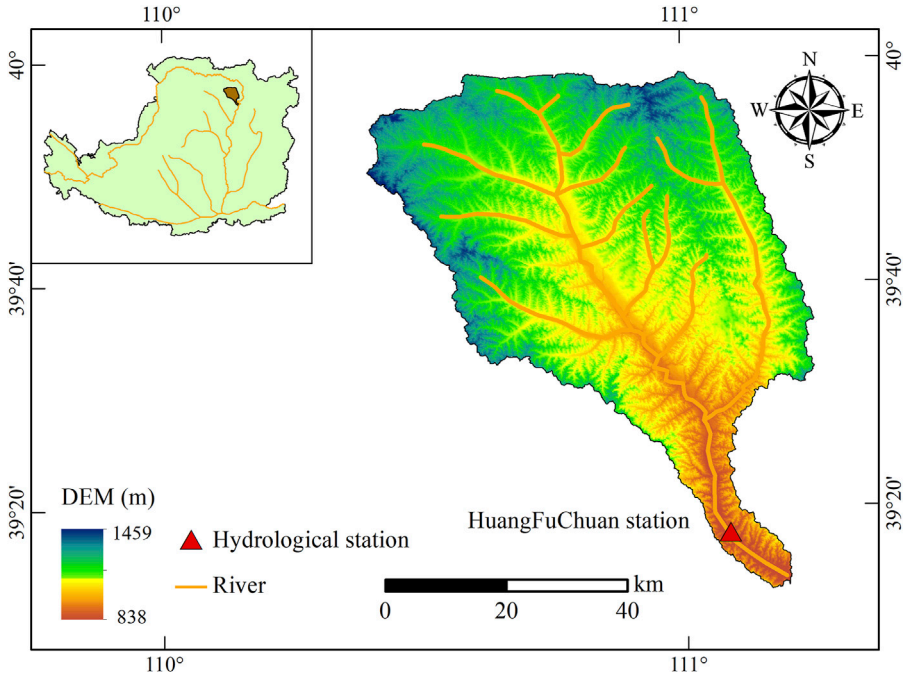


FIGURE 1
The geographical location and elevation of the Huangfuchuan Basin.

the watershed. In recent decades, the runoff of the Huangfuchuan Basin has shown a declining trend (Sui et al., 2008). Studies by experts and scholars have shown that climate change and human activities are the main reasons for the decrease in runoff and sediment movement law (Shao et al., 2014; Liu et al., 2021). However, most of the research was mainly on runoff, and few types of analysis were on the division and trend change of base-flow. In the middle reaches of the Yellow River Basin, the soil erosion control measures for a long time have resulted in the change of the flow matrix and the base-flow. Studying base-flow in the Huangfuchuan Basin is essential content for water balance and water resources management. It is also an essential basis for maintaining river ecological health (Yi and Zhou, 2017).

Based on the daily runoff data of the Huangfuchuan Basin this paper used BFI, HYSEP, and the digital filtering methods to calculate the base-flow. Through comparative analysis, this paper obtained the most suitable base-flow segmentation method in Huangfuchuan basin. This paper used the Mann-Kendall trend test, Pettitt mutation point test, Hurst index method, and wavelet analysis method to systematically research the properties of base-flow. We revealed the interannual and periodic changes of runoff and base-flow in the study area from 1960 to 2015, and predicted the future changes of baseflow. With the contradiction between the supply and demand of water resources becoming increasingly prominent, the study of this paper could provide essential information on the production mode and flow decision, the redistribution of water resources, and the optimization of the economic layout in the Huangfuchuan Basin.

Materials and methods

Study area

The Huangfuchuan Basin is located in the middle reaches of the Yellow River (Figure 1). It is in the transition zone between the Loess Plateau and the desert steppe. The latitude and longitude range are 113.3°–111.2° E, 39.2°–39.9° N, with a catchment area of 3,246 km^2 and the main steam of 137 km (Li et al., 2019). As a result of the high difference in topography and heavy rain erosion, the Huangfuchuan Basin formed a hilly gully landform. The basin has an arid and semiarid climate, with an annual average temperature of 6–7°C and an annual rainfall of 350–450 mm. The annual variation of rainfall is considerable, and most of them are concentrated in summer in the form of heavy rain. Rainfall and groundwater recharge are the primary water sources, and the correlation between annual runoff and rainfall is obvious and uneven (Hu et al., 2018). It is a typical farming-pastoral ecotone, and the land use types are mainly grassland, cultivated land, and artificial forestland. Drought and rainstorm disasters frequently occur in the Huangfuchuan Basin. Soil erosion and land desertification make the river ecological environment and hydrological process very sensitive (Wei and Jiao, 2017).

Data source

The runoff data were from the Hydrological Yearbook of the Yellow River Basin by the Yellow River Water Conservancy Commission of the Ministry of Water Resources. We used the daily runoff data of the hydrological station in the Huangfuchuan Basin from 1960 to 2015.

Base-flow segmentation method

BFI

The Base-flow index method (BFI) is a way to calculate base-flow with base-flow index as weight coefficient, which is mainly expressed by the ratio of base-flow and total runoff. There are three main steps in the calculation process. First, the same time interval was selected to divide the annual daily runoff process into 365/N periods. The study usually uses trial analysis to determine N. N could be 1, 2, 3. . ., until the base-flow index tends to be stable. This paper calculated the daily runoff data of the Huangfuchuan Basin and found that when N ≥ 5, the base-flow index is stable. So in this paper, we used parameter $N = 5$. Second, we should determine the minimum flow value of this period, and compare it with the minimum flow value of adjacent periods. Finally, we should evaluate each inflection point on the flow process line. The flow process line is connected to all inflection points. BFI method has BFI_f and BFI_k. Generally, their inflection point test factors F and K are 0.9 and 0.9792, respectively (Wels et al., 1991; Mazvimavi et al., 2004).

HYSEP

Pettyjohn and Henning first compiled the base-flow calculation program and proposed the Hydrograph-separation techniques (HYSEP) in 1979 (Pettyjonh and Henning, 1979). It mainly includes Fixed interval method, Sliding time method, and Local minimum method (Sloto and Crouse, 1996). The formula is:

$$\mathbf{N} = (2.59\boldsymbol{A})^{0.2}$$

where, A is the basin area, N is the duration of surface runoff. The area of the Huangfuchuan Basin is 3,246 km^2. Through calculation, $N = 6.09$ is obtained. The base-flow segmentation interval parameter t is an odd number that is closest to 2t and between 3 and 11. In this paper, $N = 11$ is used to calculate the base-flow.

Fixed interval method: the base-flow in the segmentation interval is equal to the minimum flow in the period. Sliding time method: the base-flow at the current moment is the minimum flow in the period before and after (2N-1)/2d. Local minimum method: in the selected time interval, the base-flow at the central moment is the minimum flow in the period before and after (2N-1)/2d. The flow at other times is obtained by linear interpolation with the flow at the major moment of the adjacent break.

Digital filtering method

The digital filtering method is a common technology of base-flow segmentation. It can decompose the signal into a high-frequency signal (surface runoff) and a low-frequency signal (base-flow) through a digital filter to achieve base-flow segmentation (Eckhardt, 2012). Digital filtering methods include Lyne-Hollick filtering (DF1) (Lyne and Hollick, 1979), Chapman-Maxwell filtering (DF2) (Chapman, 1999), Boughton-Champan filtering (DF3) (Boughton, 1993), and Eckhardt filtering (DF4) (Eckhardt, 2005).

Lyne-Hollick filtering (DF1):

$$\boldsymbol{q_t} = \partial \boldsymbol{q_{t-1}} + \frac{1+\partial}{2}(\boldsymbol{Q_t} - \boldsymbol{Q_{t-1}}),\ \boldsymbol{b_t} = \boldsymbol{Q_t} - \boldsymbol{q_t}$$

where, Q_t is the river runoff; q_t is the surface runoff; b_t is base-flow; ∂ is the filter coefficient; and when ∂ is 0.925, the effect is better.

Chapman-Maxwell filtering (DF2)

$$\boldsymbol{b_t} = \frac{\boldsymbol{k}}{2-\boldsymbol{k}}\boldsymbol{b_{t-1}} + \frac{1-\boldsymbol{k}}{2-\boldsymbol{k}}\boldsymbol{Q_t}$$

where, k is the regression coefficient, in general, $k = 0.95$.

Boughton-Champan filtering (DF3)

$$\boldsymbol{b_t} = \frac{\boldsymbol{k}}{1+\boldsymbol{C}}\boldsymbol{b_{t-1}} + \frac{\boldsymbol{C}}{1+\boldsymbol{C}}\boldsymbol{Q_t}$$

where, C is the parameter, in general, $C = 0.15$.

Eckhardt filtering (DF4)

$$\boldsymbol{b_t} = \frac{\boldsymbol{\alpha}(1-\boldsymbol{B_{max}})\boldsymbol{b_{t-1}} + (1+\boldsymbol{\alpha})\boldsymbol{B_{max}}\boldsymbol{Q_t}}{1-\boldsymbol{\alpha}\boldsymbol{B_{max}}}$$

where, B_{max} is the maximum base flow coefficient of the river.

Trend research method

Mann-Kendall trend analysis

The Mann-Kendall (MK) test is an effective tool to show the trend of change recommended by meteorological and hydrological institutions at home and abroad (Burn and Elnur, 2002). The MK test method is nonparametric. The nonparametric test method has the advantage of strictly following a specific distribution and is not affected by a few outliers. The MK test has been widely used in the change trend analysis of meteorological and hydrological time series. Its calculation formula is as follows:

Define the test statistic S:

$$S = \sum_{k=1}^{n-1} \sum_{j=K+1}^{n} sgn\left(x_j - x_k\right)$$

$$Sgn\left(X_j - X_k\right) = \begin{cases} +1 & if\ \left(X_j - X_k\right) > 0 \\ 0 & if\ \left(X_j - X_k\right) > 0 \\ -1 & if\ \left(X_j - X_k\right) > 0 \end{cases}$$

Calculate the standard statistic Z:

$$Z_{MK} = \begin{cases} \dfrac{S - 1}{\sqrt{n\ (n-1)\ (2n+5)\ /\ 18}} & for\ S > 0 \\ 0 & for\ S = 0 \\ \dfrac{S + 1}{\sqrt{n\ (n-1)(2n+5)/18}} & for\ S < 0 \end{cases}$$

where, n is the number of samples, x_j and x_k are the ranks of samples, x_i and x_j of the time series.

When Z < 0, the sequence shows an upward trend; when Z > 0, the series shows a downward trend. When the absolute value of Z is greater than or equal to 1.28, 1.64, 2.32, it means that it passes the significance test of the confidence interval of 90, 95, and 99% (Chen and Xu, 2016).

Pettitt mutation point analysis

The Pettitt mutation point test is an analysis method for the mutation point of hydrological time series based on the nonparametric test, and its premise is that there is trend change in the series (Pettitt, 1979). The formula can be calculated as follows:

$$U_t = \sum_{i=1}^{t} \sum_{j=K+1}^{T} sgn\left(x_i - x_j\right)$$

$$Sgn\left(X_j - X_k\right) = \begin{cases} +1 & if\ \left(X_i - X_j\right) > 0 \\ 0 & if\ \left(X_i - X_j\right) = 0 \\ -1 & if\ \left(X_i - X_j\right) < 0 \end{cases}$$

When testing the sequence, it uses the K-S two-sample test according to the theory of sequential statistics. The position of the mutation point (P) is the accumulation probability of the difference between the maximum values of the two-sample distribution functions.

$$K_t = max_{1 \le t \le T} |U_t|$$

$$p\left(K_t \le \alpha\right) = \left[ln\,\alpha\left(N^3 + N^3\right)/6\right]^{1/2}$$

Wavelet analysis method

Wavelet analysis provides the possibility to study time series better. It can reveal a variety of trends hidden in the time series and can qualitatively estimate the future development trend of the system (Wang, 2005).

The calculation formula of wavelet transforms:

$$W_f\ (a,b) = |a|^{-1/2} \int f(t) \times \varphi\left(\frac{t-b}{a}\right) db$$

The calculation formula of wavelet square difference:

$$V(a) = \int \left|W_f(a,b)\right|^2 db$$

where, $W_f(a,b)$ is a wavelet transform coefficient; $\varphi(t)$ is the wavelet function; $f(t)$ is the hydrologic time series; a is the scale factor of sequence period length; b is the time factor for time translation; $V(a)$ is the wavelet variance.

Hurst index

The Hurst index is an effective method to describe the long-term dependence of a time series and is used to estimate the persistence or anti-persistence of trends in time series (Bashir et al., 2020). The R/S analysis method is commonly used to obtain the Hurst index.

The Hurst index has three methods: 1) H > 0.5 means the persistence of the series, and the same trend changes in the future time series. 2) H = 0.5 stands for the randomness of the time series without persistence, indicating that the changing trend of the future time series has nothing to do with other changes. 3) H < 0.5 indicates the anti-persistence of the series, indicating that there is an opposite change trend in the future time series.

Results

Suitability analysis of base-flow segmentation

Base-flow index analysis

In this paper, the study area was segmented by nine base-flow segmentation methods in three categories (Table 1). The results of different segmentation methods differ significantly.

Among the base-flow index results calculated by HYSEP, the results calculated by LocalMin were too large, 0.249. The Fixed and Slide results were 0.179 and 0.16, respectively. BFI showed little difference and was stable at 0.1; the results obtained by the digital filtering methods were relatively scattered. In particular, the base-flow index calculated by DF2 was less than 0.1, which was the ratio of base-flow to river runoff was less than 10%. The calculation result was too small to conform to the characteristics of groundwater discharge in the study area. The other three calculation results were relatively stable, at 0.124, 0.136, and 0.113. The results of base-flow index calculated based on BFI, DF1, and DF4 are relatively close, which are all between 0.11 and 0.12. Among the nine base-flow segmentation methods, except the LocalMin, the base-flow index calculated by other methods presented a decreasing trend.

TABLE 1 The result of base-flow segmentation method.

Period	HYSEP			BFI		Digital filtering			
	Fixed	Slide	LocalMin	BFI(F)	BFI(K)	DF1	DF2	DF3	DF4
1960–1969	0.221	0.198	0.188	0.144	0.151	0.156	0.070	0.151	0.134
1970–1979	0.146	0.124	0.131	0.090	0.091	0.098	0.048	0.118	0.092
1980–1989	0.167	0.164	0.167	0.113	0.028	0.115	0.033	0.119	0.039
1990–1999	0.129	0.104	0.128	0.057	0.057	0.073	0.027	0.101	0.109
2000–2009	0.041	0.036	0.284	0.023	0.024	0.064	0.017	0.057	0.078
2010–2015	0.202	0.184	0.410	0.137	0.128	0.109	0.035	0.130	0.043
Average	0.179	0.160	0.249	0.111	0.110	0.124	0.047	0.136	0.113

TABLE 2 Characteristic plant of base-flow segmentation method.

	HYSEP			BFI		Digital filtering			
	Fixed	Slide	LocalMin	BFI (F)	BFI (K)	DF1	DF2	DF3	DF4
Mean	0.179	0.160	0.249	0.111	0.110	0.124	0.047	0.136	0.113
Variance	0.110	0.100	0.219	0.079	0.079	0.067	0.034	0.059	0.051
Maximum	0.571	0.520	0.730	0.447	0.447	0.305	0.223	0.259	0.211
Minimum	0.005	0.005	0.041	0.005	0.005	0.003	0.002	0.004	0.009
Extreme-ratio	105.498	101.933	16.770	97.037	97.037	92.171	103.051	63.870	23.600

Stability analysis

We calculated the variance and extreme-ratio (the ratio of the difference between maximum and minimum to minimum) of nine methods and compared their stability (Table 2). From the mean value, the results of the HYSEP were generally high, which was between 0.16 and 0.25; the result of BFI was relatively stable; in the digital filter methods, the values of DF1, DF3, and DF4 were between 0.1 and 0.14, and DF2 calculation had a small mean value. In terms of variance, HYSEP had high variances, and the value was above 0.1. The variance of BFI was 0.079. In the digital filter methods, DF2 and DF4 had minor variance. From the extreme-ratio, the smaller values were LocalMin in HYSEP and DF4; the largest values were Fixed in HYSEP and DF2. From the comprehensive perspective of the mean value, variance, and extreme-ratio, the variance and extreme-ratio of DF4 were relatively small. The base-flow index of different years was between 0.009 and 0.221. The calculation results of this method have slight interannual variation and are in line with the characteristics of slight fluctuation of base-flow age.

Correlation analysis

The correlation of the nine base-flow segmentation methods was calculated (Table 3). The correlation between the base-flow index values calculated by the LocalMin in the HYSEP and other methods was very low, so the following discussion did not include it.

In the HYSEP, the correlation between the Fixed and other methods was more than 0.5. Similarly, the correlation between the Slide and other methods was more than 0.5. The correlation of the base-flow index obtained by BFI (F) and BFI (K) was close to other methods, and the correlation with other methods was more than 0.45. In the digital filtering methods, the correlation coefficient between DF1 and other methods was between 0.57 and 0.77, and DF2 and other methods were between 0.46 and 0.66. The correlation coefficient between DF3 and other techniques was 0.63–0.89. The correlation between F4 and other methods was higher than 0.77 except for the correlation with DF2. In the three categories of methods, the correlation between the BFI and HYSEP is better than digital filtering methods, indicating that BFI is close to the base-flow segmentation results of HYSEP.

By analyzing the base-flow segmentation method, the value of the LocalMin in HYSEP was too small, and DF2 in the digital filtering methods was too high; neither of these two methods can accord with the change characteristics of base-flow in the study area. In general, the base-flow index of DF4 had the highest correlation with the other methods. Its variance and extreme-ratio were generally very small, showing high stability. In the Huangfuchuan Basin, the base-flow result calculated by DF4 in the digital filtering method was more appropriate, and base-flow results were also used in this paper to analyze the base-flow changes.

TABLE 3 Correlation of base-flow segmentation methods.

	HYSEP			BFI		Digital filtering			
	Fixed	Slide	LocalMin	BFI(F)	BFI(K)	DF1	DF2	DF3	DF4
Fixed	1.000								
Slide	0.961	1.000							
LocalMin	0.157	0.181	1.000						
BFI(F)	0.905	0.949	0.180	1.000					
BFI(K)	0.891	0.927	0.171	0.985	1.000				
DF1	0.706	0.752	0.348	0.711	0.701	1.000			
DF2	0.513	0.551	0.042	0.469	0.492	0.576	1.000		
DF3	0.862	0.851	0.030	0.773	0.756	0.637	0.657	1.000	
DF4	0.879	0.905	0.061	0.843	0.829	0.774	0.629	0.891	1.000

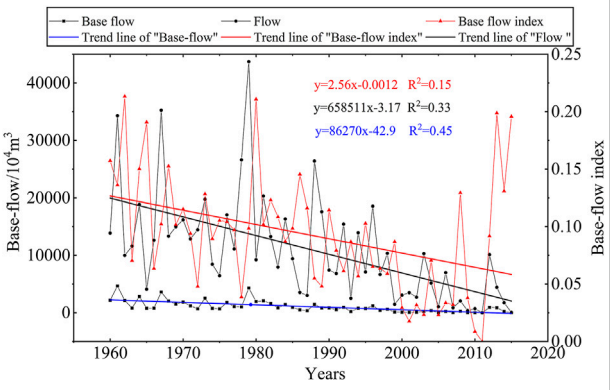


FIGURE 2
Change curves of runoff, base-flow, and base-flow index.

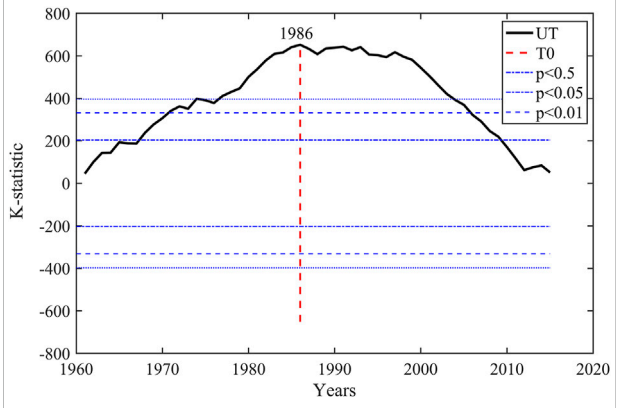


FIGURE 3
Mutagenicity analysis of the Huangfuchuan Basin.

Analysis of base-flow characteristics

Trend analysis

Through the trend line in the research area, the runoff presented a decreasing trend (Figure 2). The maximum runoff occurred in 1979, which was 43,712.22 × 10^4 m^3, and the annual average runoff was 1,100.27 × 10^4 m^3. The base-flow index also showed a downward trend. The maximum base-flow index was 0.21, which appeared in 1962. The average base-flow was 10,578.35 × 10^4 m^3, the maximum yearly base-flow was 46.679 × 10^4 m^3. The Mann-Kendall trend approach to calculate the Z-testing statistical value was −6.349, indicating that the base-flow presented a significant downward trend and reached a significant level of 99% reliability.

Mutagenicity analysis and persistent

We used the Pettitt mutation point test method to analyze the base-flow mutation years in the Huangfuchuan Basin from 1960 to 2015 (Figure 3). The mutation year was 1986, and the confidence level was 99%. The Hurst index of base-flow in the study area was 0.84 (H = 0.84). H > 0.5, there was a long-term correlation between the future trend and its historical trend. As seen from the above, the base-flow in the study area is decreasing continuously, and it will be further reduced in the future.

Periodicity

The periodic variation of base-flow in the study area was analyzed by a wavelet coefficient diagram and wavelet variance (Figure 4). The base-flow presented periodic variation characteristics at scales of 3–6, 7–18, and 19–32. In the 19–32 years time scale, three oscillations alternated between abundant and dry. The center of the periodic change was 31 years, corresponding to the first peak value of the wavelet variance. The oscillation was weak on the time scale of 3–6 years and gradually disappeared in 1980. On a time scale of 7–18 years, there were six oscillations and the oscillations gradually became weaker.

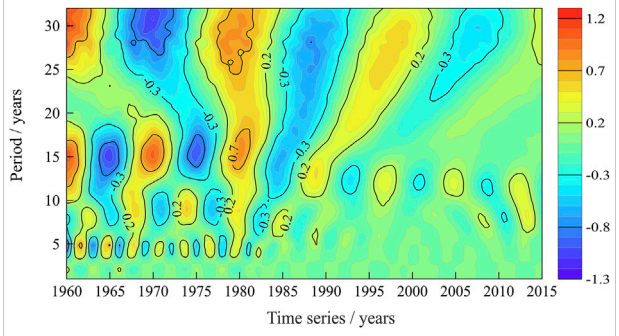


FIGURE 4
The wavelet coefficient of base-flow in the Huangfuchuan Basin.

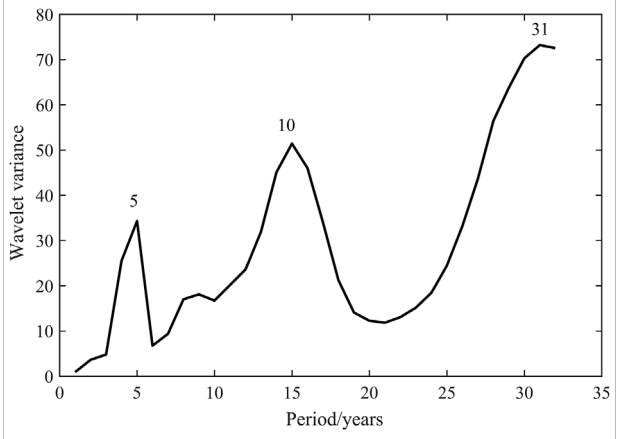


FIGURE 5
The wavelet variance of base-flow in the Huangfuchuan Basin.

In the wavelet variance of base-flow, there are three peaks, corresponding to 5, 15, and 31 years (Figure 5). Below 10 years, the cycle oscillates violently without apparent regularity. On scales 15 and above, periodic and regular regions were evident. The first peak value of the wavelet square difference was 31 years, which was the first main period of base-flow change.

Discussion

The Huangfuchuan Basin is one of the areas with the most severe soil erosion on the Loess Plateau. In recent years, climate change and human activities have changed the base-flow, so researching the base-flow is indispensable for constructing the local ecological environment. This paper used HYSEP, BFI, and the digital filters methods to calculate the base-flow in the study area. By comparing the stability and correlation of the nine methods, the variance and extreme-ratio of DF4 in the digital filter methods were generally minor, and the correlation with other methods was also high. It is concluded that the base-flow segmentation result of DF4 was the most appropriate. According to the DF4 results, the average annual base-flow in the study area is $1{,}057.84 \times 10^4$ m^3, and the average yearly base-flow index is 0.113. Liu et al used BFI and the digital filtering methods to calculate the base-flow of the Huangfuchuan Basin, and the study showed the changes of runoff and base-flow showed a decreasing trend, which was consistent with the results of this paper (Liu et al., 2016). In addition, this paper analyzed the periodicity, mutation year and future trend of base-flow. The abrupt change of base-flow occurred in 1986. The first principal period of the base-flow is 31 years. There was also a trend of further decrease in the future.

Base-flow segmentation is a common problem in hydrological analysis and calculation of watersheds. At present, there are many methods for base-flow segmentation; however, there is no universally accepted technology. The segmentation results of each technology are not exactly the same, and sometimes even very different. Due to the differences in hydrogeological conditions and basic abortion flow processes, the appropriate methods are different in different basins. In the Yangtze River Basin, BFI is the most suitable method for base-flow segmentation, while in the Fenhe River Basin and the Huangfuchuan Basin, digital filtering method is the most suitable method (Wang et al., 2015; Lv et al., 2021). In addition to the above automatic segmentation methods, hydrological modeling methods also began to develop rapidly. SWAT was used to study a small area in the United States and the Luohe River basin, and the study showed that the accuracy of SWAT was higher than that of the digital filtering method (Yang et al., 2003).

Base-flow is a component of runoff, which is affected by climatic characteristics, precipitation, human activities, vegetation and soil characteristics (Hu et al., 2021). Li et al showed that forest disturbance not only increased the runoff of the basin, but also increased the base-flow. Climate change, contrary to the effect of forest disturbance, reduced the flow of all runoff components (Li et al., 2018). Guo et al showed that different land use types, precipitation, drainage density and river gradient affected the change of base-flow on the Loess Plateau. The change of runoff and base-flow shows a decreasing tendency in the Huangfuchuan Basin (Guo et al., 2011). According to the climate characteristics of the Huangfuchuan Basin, the temperature shows a rising trend, while the precipitation shows a declining trend (Hu et al., 2018). Evapotranspiration intensifies the water pressure, and the decrease of rainfall caused by climate change has an important impact on runoff and base-flow. In addition, the changes of runoff and base-flow are also affected by human activities, such as conversion of farmland to forest and water conservation processes, which change the underlying surface and reduce the runoff (Zhao et al., 2013; Liu et al., 2016). In recent years, there has been excessive exploitation of coal resources and the increase of urban land area in the Huangfuchuan Basin (Jin et al., 2013; Yu et al., 2014). These

human activities have greatly reduced the water area and affected runoff and base-flow.

Due to the limitations of data, research methods, and techniques, this paper still has some deficiencies and uncertainty. In this paper, we used a variety of methods widely recognized to calculate the base-flow and selected the most suitable segmentation method. However, base-flow is difficult to measure directly, so it was difficult to verify the results of base-flow segmentation method. The results in this paper were only the estimated value of the actual base-flow, rather than the real value (Lei et al., 2021). Using the results to analyze the variation characteristics of base-flow, and only using the method of mathematical statistics, there are certain errors between the results and the actual situation. Therefore, in future research, we should further optimize the practicability of the base-flow segmentation method in the Huangfuchuan Basin.

Conclusion

In this paper, nine base-flow segmentation methods were used to calculate the base-flow, and then their suitability in the Huangfuchuan Basin was discussed. Based on MK trend analysis, Pettitt mutation point test, Wavelet analysis method, and Hurst index, the characteristics of base-flow were studied. The main conclusions of this paper are as follows:

1) Among the results of HYSEP, BFI, and the digital filtering methods for base-flow, the variance and extreme-ratio of DF4 in the digital filtering methods were slight. The correlation between DF4 and other base-flow segmentation methods was the highest. The DF4 results were consistent with the stable characteristics of base-flow; thus, it was the most stable and appropriate method in the Huangfuchuan Basin.
2) The runoff and base-flow presented a decreasing trend. We used MK trend test to analyze the movement of base-flow, which showed a decreasing trend, and the annual base-flow was $1{,}057.84 \times 10^4$ m^3.
3) The Pettitt abrupt change point and the Hurst index were used to analyze the catastrophe point and future change trend of base-flow. The abrupt change point occurred in 1986. The trend of future development was positively correlated with the past change, showing a trend of continuous decrease.
4) According to the results of wavelet analysis, the base-flow showed a periodic change of 3–6, 7–18, 19–32 scales. Base-flow showed a period change centered on 31 years, corresponding to the first front value of wavelet square.

Data availability statement

The original contributions presented in the study are included in the article/supplementary materials, further inquiries can be directed to the corresponding author.

Author contributions

Conceptualization—YH and XM; methodology—YH, YZ, LJ, and YL; Writing—original draft preparation, YZ; Writing—review and editing, YH and XM. All authors have read and agreed to the published version of the manuscript.

Funding

This research was jointly supported by the National Natural Science Foundation of China (42077075, 41671285, U2243211), the China Postdoctoral Science Foundation (Grant No. 2018M633602), Postdoctoral Research Fund of Shaanxi Province (Grant No. 2017BSHEDZZ144), and the Natural Science Basic Research Plan in Shaanxi Province of China (Grant No. 2021JQ-449).

Acknowledgments

We would like to thank the reviewers and editors for their valuable comments and suggestions.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Arnold, J. G., Allen, P. M., Muttiah, R. S., and Bernhardt, G. (1995). Automated base flow separation and recession analysis techniques. *Ground Water* 33, 1010–1018. doi:10.1111/j.1745-6584.1995.tb00046.x

Bashir, B., Cao, C., Naeem, S., Joharestani, M., Mumtaz, F., Afzal, H., et al. (2020). Spatio-temporal vegetation dynamic and persistence under climatic and anthropogenic factors. *Remote Sens.* 12, 2612. doi:10.3390/rs12162612

Bastin, L., Gorelick, N., Saura, S., Bertzky, B., Dubois, G., Fortin, M. J., et al. (2019). Inland surface waters in protected areas globally: Current coverage and 30-year trends. *PLoS One* 1, e0210496. doi:10.1371/journal.Pone.0210496

Boughton, W. C. (1993). *A hydrograph-based model for estimating the water yield of ungauged catchments*. Australia: Hydrology and Water Resources Symposium.

Brutsaert, W., and Nieber, J. L. (1977). Regionalized drought flow hydrographs from a mature glaciated plateau. *Water Resour. Res.* 13, 637–643. doi:10.1029/WR013i003p00637

Burn, D. H., and Elnur, M. (2002). Detection of hydrologic trends and variability. *J. Hydrology* 255, 107–122. doi:10.1016/S0022-1694(01)00514-5

Chang, M. Q. (2009). The securing of water resource issue and its nature. *China Saf. Sci. Journal(CSSJ)* 19, 1003–3033. doi:10.16265/j.cnki.issn1003-3033.2009.09.027

Chapman, T. (1999). A comparison of algorithms for stream flow recession and baseflow separation. *Hydrol. Process* 13, 701–714. doi:10.1002/(SICI)1099-1085(19990415)13:5<701::AID-HYP774>3.0.CO;2-2

Chapman, T. G. (1991). Comment on "evaluation of automated techniques for base flow and recession analyses" by R. J. Nathan and T. A. McMahon. *Water Resour. Res.* 27, 1783–1784. doi:10.1029/91WR01007

Chen, L. Q., Liu, C. M., and Li, F. D. (2006). Review of base-flow research. *Prog. Geogr.* 25, 1–15. doi:10.1007/s00198-012-2119-2

Chen, Z. P., and Xu, Q. (2016). Analysis of precipitation characteristics in Jinhua by Mann-Kendall test method. *Bull. Sci. Technol.* 32, 47–50. doi:10.13774/j.cnki.kjtb.2016.06.012

Eckhardt, K. (2008). A comparison of baseflow indices, which were calculated with seven different baseflow separation methods. *J. Hydrology* 352, 168–173. doi:10.1016/j.jhydrol.2008.01.005

Eckhardt, K. (2005). How to construct recursive digital filters for baseflow separation. *Hydrol. Process.* 19, 507–515. doi:10.1002/hyp.5675

Eckhardt, K. (2012). Technical note: Analytical sensitivity analysis of a two parameter recursive digital baseflow separation filter. *Hydrol. Earth Syst. Sci.* 16, 451–455. doi:10.5194/hess-16-451-2012

Guo, J. T., Zhang, Z. Q., Wang, S. P., Zhang, J. J., and Peter, S. (2011). Features of baseflow and its influencing factors for small watersheds in Loess Hilly and Gully region. *Bull. Soil andWater Conservation* 31 (1), 87–92. doi:10.13961/j.cnki.stbctb.2011.01.024

Hu, Y. N., Duan, W. L., Chen, Y. N., Zou, S., Kayumba, P. M., and Sahu, N. (2021). An integrated assessment of runoff dynamics in the Amu Darya River Basin: Confronting climate change and multiple human activities, 1960–2017. *J. Hydrology* 603, 126905. doi:10.1016/j.jhydrol.2021.126905

Hu, Z. D., Xia, T., Luo, L., and Wang, Z. (2018). Impact from climate change on amount of river runoff - a case of Huangfuchuan Watershed. *Water Resour. Hydropower Eng.* 49, 30–36. doi:10.13928/j.cnki.wrahe.2018.02.005

Janke, B. D., Finlay, J. C., Hobbie, S. E., Baker, L. A., Wilson, B. N., Nidzgorski, D., et al. (2014). Contrasting influences of stormflow and baseflow pathways on nitrogen and phosphorus export from an urban watershed. *Biogeochemistry* 121, 209–228. doi:10.1007/s10533-013-9926-1

Jin, S. Y., Lin, C. Y., Zhang, Z., and Chen, Y. (2013). Effect of underlying surface change on"July 21"Flood in Huangfuchuan Basin. *Yellow River* 35 (04), 19–21. doi:10.3969/j.issn.1000-1379.2013.04.008

Lei, Y., Jiang, X., Geng, W., Zhang, J. Y., Zhao, H., Ren, L. Q., et al. (2021). The variation characteristics and influencing factors of base flow of the Hexi Inland Rivers. *Atmosphere* 12, 356. doi:10.3390/ATMOS12030356

Li, C. Q., Guo, S. L., Zhang, J., and Lin, K. R. (2009). A modified TOPMODEL based on horton infiltration capacity curve. *J. China Hydrology* 29, 4–7. doi:10.1042/BSR20080061

Li, M., Zhang, H. L., and Meng, C. C. (2019). Study on characteristics of water-sediment relationship and key influencing factors in Huangfuchuan Watershed of Yellow River. *Adv. Sci. Technol. Water Resour.* 39, 27–35. CNKI:SUN:SLSD.0.2019-05-006.

Li, Q., Wei, X. H., Zhang, M. F., and Liu, W. F. (2018). The cumulative effects of forest disturbance and climate variability on streamflow components in a large forest-dominated watershed. *J. Hydrology* 557, 448–459. doi:10.1016/j.jhydrol.2017.12.056

Liao, Z. M., Li, Y. Y., Xiong, W. S., Wang, X., Dan, L., and Zhang, Y. L. (2020). An in-depth assessment of water resource responses to regional development policies using hydrological variation analysis and system dynamics modeling. *Sustainability* 12, 1–18. doi:10.3390/su12145814

Lin, Q. C., and Li, H. E. (2010). Influence and guarantee on ecological basic flow of Weihe River from Baojixia water diversion. *J. Arid Land Resour. Environ.* 24, 114–119. doi:10.1631/jzus.A1000244

Liu, B. J., Lei, X. H., Liu, Z. S., Quan, J., and Wang, H. (2016). Respective characteristics and mutual response relationship of precipitation and base-flow in Huangfuchuan Basin. *Yellow River* 38 (4), 7–12. doi:10.3969/j.issn.1000-1379.2016.04.002

Liu, B. L., Li, P., Yang, Y. Y., Ren, Z. P., and Long, S. B. (2021). Response of variation of runoff and sediment to landscape pattern in Huangfuchuan Basin. *Res. Soil Water Conservation* 28, 102–107. doi:10.13869/j.cnki.rswc.2021.05.012

Lv, X. L., Ji, S. B., and Chou, Y. Q. (2021). Analysis of characteristics of base flow of Fen River upstream and its influencing factor. *Water Conserv.* 11, 1–12. doi:10.3880/j.issn.1004-6933.2022.02.020

Lyne, V. D., and Hollick, M. (1979). "Stochastic time-variable rainfall-runoff modeling," Institute of Engineers Australia National Conference, 89–93.

Mazvimavi, D., Meijerink, M. J., and Stein, A. (2004). Prediction of base flows from basin characteristics: a case study from Zimbabwe/prévision de débits de base à partir de caractéristiques du bassin: une étude de cas au Zimbabwe. *Hydrological Sci. J.* 49, 703–715. doi:10.1623/hysj.49.4.703.54428

Partington, D., Brunner, P., Simmons, C. T., Werner, A. D., Therrien, R., and Maier, H. R. (2012). Evaluation of outputs from automated baseflow separation methods against simulated baseflow from a physically based, surface water-groundwater flow model. *J. Hydrology* 458, 28–39. doi:10.1016/j.jhydrol.2012.06.029

Pettitt, A. N. (1979). A non-parametric approach to the change-point problem. *Appl. Stat.* 28, 126. doi:10.2307/2346729

Pettyjonh, W. A., and Henning, R. J. (1979). *Preliminiary estimate of ground-water recharge rates, related stream-flower and water quality in Ohio*. Columbus, Ohio, United States: Ohio State University Water Resources Center Project Completion Report.

Rutledge, A. T. (1993). *Computer programs for describing the recession of ground-water discharge and for estimating mean ground-water recharge and discharge from streamflow records*. United States: US Geological Survey Water-Resources Investigations Report.

Santhi, C., Allen, P. M., Muttiah, R. S., Arnold, J. G., and Tuppad, P. (2008). Regional estimation of base flow for the conterminous United States by hydrologic landscape regions. *J. Hydrology* 351, 139–153. doi:10.1016/j.jhydrol.2007.12.018

Shao, G. W., Guan, Y. Q., Guan, Z. Q., Sha, Z. G., and Zhang, D. R. (2014). Analysis of variation trend of runoff and driving factors in Huangfuchuan. *J. Water Resour. Water Eng.* 25, 54–56. doi:10.11705/j.issn.1672-643X.2014.03.11

Sloto, R. A., and Crouse, M. Y. (1996). *HYSEP: a computer program for streamflow hydrograph separation and analysis*. United States: US Geological Survey Water-Resources Investigations Report.

Sui, J., He, Y., and Karney, B. W. (2008). Flow and high sediment yield from the Huangfuchuan watershed. *Int. J. Environ. Sci. Technol. (Tehran).* 5, 149–160. doi:10.1007/BF03326008

Tallaksen, L. M. (1995). A review of baseflow recession analysis. *J. Hydrology* 165, 349–370. doi:10.1016/0022-1694(94)02540-R

Veldkamp, T., Wada, Y., Aerts, J., Doll, P., Gosling, S. N., Liu, J., et al. (2017). Water scarcity hotspots travel downstream due to human interventions in the 20th and 21st century. *Nat. Commun.* 8, 15697. doi:10.1038/ncomms15697

Wang, G., Lu, C. P., Li, S. L., Chen, S., and Li, J. J. (2015). Applied comparision of five separation methods of base flow at Luoshan station of Yangtze River. *J. Water Resour. Water Eng.* 26, 118–123. doi:10.11705/j.issn.1672-643X.2015.03.25

Wang, W. S. (2005). *Hydrologic wavelet analysis*. Beijing: Chemical Industry Press.

Wei, Y. H., and Jiao, J. Y. (2017). Variation tendency and periodic characteristics of streamflow and sediment discharge in Huangfuchuan Watershed from 1955 to 2013. *Res. Soil Water Conservation* 24, 1–6. doi:10.13869/j.cnki.rswc.2017.03.001

Wels, C., Cornett, R. J., and Lazerte, B. D. (1991). Hydrograph separation: A comparison of geochemical and isotopic tracers. *J. Hydrology* 122, 253–274. doi:10.1016/0022-1694(91)90181-G

Wittenberg, H., and Sivapalan, M. (1999). Watershed groundwater balance estimation using streamflow recession analysis and baseflow separation. *J. hydrology* 219, 20–33. doi:10.1016/s0022-1694(99)00040-2

Xu, R. H., Wang, X. G., and Zheng, W. (2016). Research progresses in Baseflow separation methods. *Bull. Soil Water Conservation* 36, 352–359. doi:10.13961/j.cnki.stbctb.2016.05.061

Yang, G., Hao, F. H., and Liu, C. (2003). The study on Baseflow estimation and evaluation based on SWAT-Luohe River Basin as Example. *Prog. Geogr.* 22, 463–471. doi:10.1002/ppap.201400126

Yi, L. Q., and Zhou, R. (2017). An analysis of the various characteristics and influencing factors of the settlement in the view of new district construction of city in recent 10 years. *J. Inn. Mong. Normal Univ.* 46, 140–145. CNKI:SUN:NMGS.0.2017-01-029.

Yu, F., Li, X. B., and Wang, H. (2014). Optimization of land use pattern based on eco-security: a case study in the huangfuchuan watershed. *Acta eco. Sin.* 34 (12), 3198–3210. doi:10.5846/stxb201307011807

Zhang, X. S., Srinivasan, R., Arnold, J., Lzaurralde, C. R., and Bosch, D. (2011). Simultaneous calibration of surface flow and baseflow simulations: a revisit of the SWAT model calibration framework. *Hydrol. Process.* 25, 2313–2320. doi:10.1002/hyp.8058

Zhao, G. J., Mu, X. M., Wen, Z. M., Wang, F., and Gao, P. (2013). Impacts of precipitation and human activities on streamflow and sediment load in the Huangfuchuan Watershed. *Sci. Soil Water Conservation* 11 (04), 1–8. doi:10.16843/j.sswc.2013.04.001

Zuo, H. F., Wu, S. L., Shao, J. L., and Cui, Y. L. (2007). Application of computerized base-flow separation method with BFI program in mountainous areas. *J. China Hydrology* 27, 69–71. doi:10.3969/j.issn.1000-0852.2007.01.018

frontiers | Frontiers in Environmental Science

ORIGINAL RESEARCH
published: 19 August 2022
doi: 10.3389/fenvs.2022.927765

Characterization and Risk Assessment of Heavy Metals in Surface Sediments From Jian and Moyang Rivers in Western Guangdong

*Jie Feng, Yuemin Lin, Mingkun Li, Tingping Ouyang and Mingjie Yu**

School of Geographic Sciences, South China Normal University, Guangzhou, China

The river environment is complex and receives a variety of contaminants from numerous sources that are persistent, bioaccumulative, and toxic. The distribution, source, contamination, and ecological risk status of Zn, Pb, Cu, Ni, Cr, and Cd were evaluated in the surface sediments at 45 sites on the Moyang and Jian rivers in Western Guangdong, China. Single pollution indices, including contamination factor (CF) and enrichment factor (EF), revealed that Zn, Pb, Ni, Cu, and Cd showed moderate to significant enrichment. To overcome the limitation of the single element indices, a range of sediment quality indices, including modified contamination index (mCd), pollution index (PI), and modified pollution index (MPI), were utilized to ascertain the sediment quality. The sediment in the study area is deemed to be slightly to extremely polluted. The sediment quality guidelines (SQGs), potential ecological risk index (RI), and modified ecological risk index (MRI) were used to assess possible ecological risks. According to the SQGs, Pb, Ni, and Cu have the potential to induce biological effects. The RI indicated that the sediment poses a low ecological risk. However, the MRI indicated that the ecological risk of the sediment was moderate to very high. The accuracy of the single and multi-element indices and ecological risk assessment were evaluated in the river using the principal component analysis (PCA) and cluster analysis (CA), showing an anthropogenic impact. Results demonstrate the need to pay attention to the ecological environment of small rivers, which are sensitive to their surroundings.

Keywords: surface sediment quality, pollution indices, modified ecological risk index, principal component analysis, cluster analysis, spatial distribution, Jian and Moyang rivers, Western Guangdong

OPEN ACCESS

Edited by:
Zhenzhong Zeng, Southern University of Science and Technology, China

Reviewed by:
Venkatramanan Senapathi, Alagappa University, India
Omowunmi H. Fred-Ahmadu, Covenant University, Nigeria

****Correspondence:***
Mingjie Yu
yumj@m.scnu.edu.cn

Specialty section:
This article was submitted to Freshwater Science, a section of the journal Frontiers in Environmental Science

Received: *25 April 2022*
Accepted: *16 June 2022*
Published: *19 August 2022*

Citation:
Feng J, Lin Y, Li M, Ouyang T and Yu M (2022) Characterization and Risk Assessment of Heavy Metals in Surface Sediments From Jian and Moyang Rivers in Western Guangdong. Front. Environ. Sci. 10:927765. doi: 10.3389/fenvs.2022.927765

1 INTRODUCTION

Metal pollutants pose potential harmful effects on human health and the whole ecosystem due to their inherent toxicity, persistence, non-degradability, and bioaccumulation (Wei et al., 2016). Metals in aquatic environments originate from natural sources (mainly weathering of soil and rock, erosion, forest fires, and volcanic eruptions) and anthropogenic activities (industrial effluents, mining and refining, agricultural drainage, domestic discharges, and atmospheric deposition) (Karbassi et al., 2007; Malik et al., 2009; Davutluoglu et al., 2011). Toxic heavy metal elements (such as Pb, Hg, and Cr) enter the human body through the food chain, causing serious harm to the human body (Liu et al., 2018).

Sediment acts as an ultimate receptor of pollutants and a potential secondary source of overlying water. The pollutants from residential sewage and agricultural water may be stored in the sediments

through hydrodynamic and biogeochemical processes (Wu et al., 2017). The harm caused by heavy metal sediments is mainly reflected in the "secondary pollution". In addition, heavy metals accumulated in sediments enter the water due to changes in interface environmental conditions such as pore water pH, decomposition of organic matter, biological activities, storms and dumping of port dredged materials (Tang et al., 2015). Therefore, the analysis of river sediments is a useful approach to characterize pollution in an ecosystem (Akcay et al., 2003).

A number of indices have been developed to accurately assess the metal contamination in sediments and its ecological risk. The common indicators mainly include contamination factor (CF), enrichment factor (EF), (modified) degree of contamination (Cd or mCd), (modified) pollution index (PI or MPI), sediment quality guidelines (SQGs), and (modified) potential ecological risk index (RI or MRI) (Yuan et al., 2014; Vaezi et al., 2015; Duodu et al., 2016; Kumar et al., 2018). These indicators can be divided into two broad categories. SQGs, RI, and MRI indicate the risk of sediments based on the metal concentrations, while CF, EF, Cd, mCd, PI, and MPI denote the enrichment compared with the background concentrations based on the total metal concentrations (Yuan X et al., 2014). The CF, EF, Cd, and mCd assess the sediment quality by using single pollution indices. Meanwhile, PI and MPI assess the sediment quality by using multi-element indices (Duodu et al., 2016). MRI and MPI can normalize the impact of terrestrial sedimentary inputs using EFs (Brady et al., 2015; Duodu et al., 2016). Numerous studies have used more than one index to have a more comprehensive analysis of the contamination in sediment (Vaezi et al., 2015; Kumar et al., 2018; Siddiqui and Pandey, 2019).

Recently, urban-industrial-driven economic development has dramatically altered the aquatic environments with increasingly high input of metals and other pollutants. Terrestrially derived metals, whether geogenic or anthropogenic, flow into rivers and accumulate in sediments (Omwene et al., 2018). Studies on heavy metal pollution in river sediments have mostly focused on large rivers, such as the Yangtze and Pearl rivers in China (Fu et al., 2013; Zhao et al., 2016), the Korotoa and Buriganga rivers in Bangladesh (Mohiuddin et al., 2016), and the Tajum River in Indonesia (Budianta, 2020). However, studies on small rivers are scarce. The Pearl River is located in the developed Pearl River Delta region of Guangdong Province, where sediment pollution has received considerable attention (Zhao et al., 2016). However, few investigations have been conducted on the pollution status of sediments in the Jian and Moyang rivers in Western Guangdong's underdeveloped areas.

Although the economic and social development in Western Guangdong is slow, it is the main agricultural contribution area in Guangdong. In addition, with the transfer of industries in the Pearl River Delta, the industrial domestic pollution of Western Guangdong has increased (Zhong and Yang, 2001; Chen, 2014). Therefore, the pollution problems of small rivers located in Western Guangdong have frequently occurred in recent years, which need to be paid attention to.

In this study, 45 surface river sediment samples collected from Moyang and Jian rivers in Western Guangdong, China, were studied. This study aimed to use sediment quality indicators, such as CF, EF, PI, MPI, RI, MRI, and SQGs to examine the contamination and ecological status of the Moyang and Jian Rivers sediment. Moreover, this study sought to identify natural and anthropogenic sources of metal contamination based on multivariate statistical analysis. To the best of our knowledge, this work is the first to study heavy metal contamination and bioavailability for the ecological risks in the Moyang and Jian rivers, which might construct a more susceptible and accurate evaluation system for similar river ecosystems in China.

2 MATERIALS AND METHODS

2.1 Study Area

The study area (Western Guangdong, 109°31′E to 112°21′E and 20°13′N to 22°41′ N) is located in the western part of Guangdong Province. This area has a subtropical climate, which is affected by the maritime monsoon and subtropical high-pressure climate. The main source of surface runoff is rainfall. Due to the uneven distribution of rainfall over time, annual runoff varies greatly. Flood season is mainly from April to September, and non-flood season is from October to March of the next year. The coastal rivers in the study area include Jian, Moyang, Jiuzhou, Nandu, and Suixi Rivers. Our study focused on sediment pollution in the Jian and Moyang rivers, which are longer rivers in Western Guangdong, with a total length of 231 and 199 km, respectively. The urban land accounts for about 10% of the province in Western Guangdong, while the arable land accounts for 32.4% of the province. The economic structure of the study area accounts for a large proportion of agricultural output value (Zhang, 2020). Therefore, the environmental pollution caused by pesticides, fertilizers, and livestock is obvious during agricultural production. The study area, as the undertaking area for industrial transfer in the Pearl River Delta, has experienced significant growth in economic and social development in recent years (Zhang, 2013; Huang et al., 2020), which probably increased industrial and urban life pollution.

2.2 Experimental Design

Considering the distribution of land utilization types, dam distribution, and flood season time, 45 surface sediment samples (0–10 cm) were collected using corers from November 1st to 7th, 2020 (**Figure 1**). It was found that periods of low flows in dry seasons lead to higher concentrations of heavy metals in channel-bed sediments, whereas wet seasons are characterized by a lower metal content in the bed and a higher metal content in suspended sediments (Gaiero et al., 1997; He et al., 1997). In order to avoid the dilution of heavy metals in river sediments during flood season, we sampled river sediments during non-flood season. A total of 45 sampling sites were set up downstream of Jian River (S1-S20) and downstream of Moyang River (S20-S45). They are distributed in agricultural areas, urban areas, estuaries, and other types of land areas. Reservoirs and dams weakened the hydrodynamic intensity of the river and accelerated

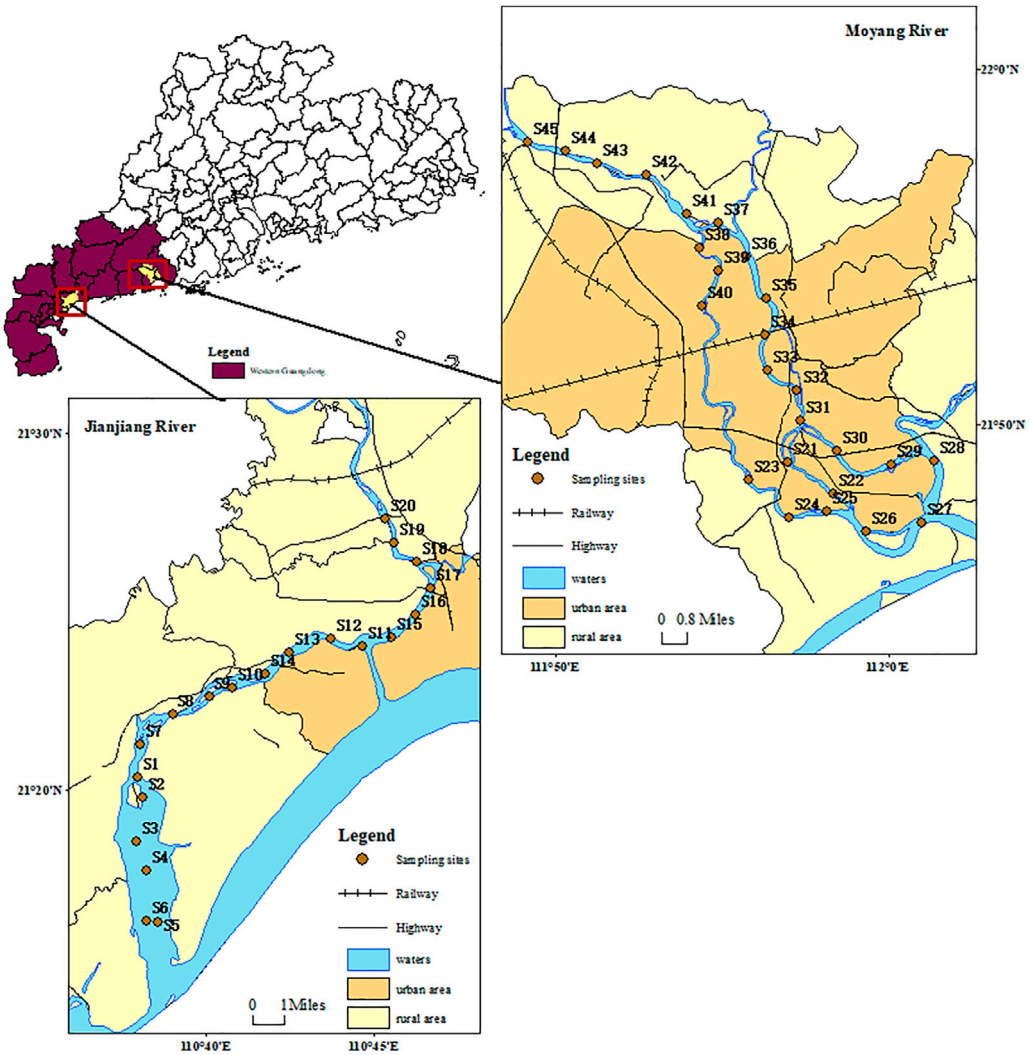


FIGURE 1 | Locations of sampling sites of surface sediments in Jian and Moyang river basins.

the deposition of suspended solids (Kang et al., 2021). But there is no reservoir and dam in the section where the sampling site is located. Locations of sampling stations were determined using Global Positioning System (GPS) (**Supplementary Table S1**). The collected samples were collected in polyethylene sealed bags and transported to the laboratory of Guangdong Polytechnic of Water Resources and Electric Engineering for analysis.

2.3 Sampling and Analytical Methods

After the sediment samples were freeze-dried, they were ground through a 200-mesh sieve and stored for testing. The analytical method of heavy metal concentration in sediment was as follows: 0.5 g of sediment was taken into a digestion tube. Then, 3–5drops of pure water, 8 ml of concentrated nitric acid, 5 ml of hydrofluoric acid, and 5 ml of hydrogen peroxide were added in sequence. Then, a graphite digester (GDI-20, Guangzhou Jichuang Instrument Co., Ltd. China) was used to digest until 1–2 ml of sample solution remained in the tube. The solution was filtered into a 50 ml colorimetric tube with a 0.45 um water-based disposable filter needle. the volume was diluted to 50 ml with pure water for testing. The prepared samples were analyzed by atomic absorption spectrometry (AAS). Seven metals were analyzed in the sediment sample, which are Cu, Pb, Zn, Cd, Cr, Ni, Fe, and Mn. The quality control process of heavy metal analysis in sediments was as follows: The reagents used in the experimental analysis were all high-grade samples. The relative standard deviation (RSD) of experiments was less than 5%.

2.4 Data Analysis

The average of the metal concentrations was calculated in this study. The data were analyzed using the multivariate statistical tools, principal component analysis (PCA) and cluster analysis (CA), to group variables based on similarities and their sources. Single pollution indices, multi-element indices, and ecological risk assessment were combined to analyze heavy metal pollution in the study area. Given that the study area is located in the South

China–Youjiang water system and the low mountain and hilly water system, the evaluation method in this study took the average background value of these water systems as the evaluation standard C_b, Zn: 67.60 mg/kg, Pb: 28.49 mg/kg, Ni: 20.13 mg/kg, Mn: 575.35 mg/kg, Cu: 19.00 mg/kg, Cr: 51.30 mg/kg, and Cd: 0.1298 mg/kg (Shi et al., 2016).

2.4.1 Single Element Indices

Contamination factor (CF) provides the ratio of an element at background sites, a reference value or a national criterion for that metal (Duodu et al., 2016). This indicator provides information about how an element has been concentrated at the site of interest relative to a background site, which indicates the input of metals by human activities (Ahmed et al., 2016). **Eq. 1** shows how CF is calculated (Hakanson, 1980).

$$CF = \frac{C_i}{C_b}, \tag{1}$$

where CF is the contamination factor, C_i is the concentration of interest metal at a site, and C_b is the concentration of the same metal at a background or reference site. Four qualitative terminologies are used to describe the CF by Hakanson (1980): CF < 1: low contamination; 1 ≤CF < 3: moderate contamination; 3 ≤ CF < 6: considerate contamination; CF ≥ 6: very high contamination.

EF is normalized against an element, which compares the ratio of the element of interest to a "conservative element" in a given sample to the same ratio in a local background (Brady et al., 2015; Islam et al., 2018). "Conservative element" commonly uses the following elements: Al, Fe, Mn, Sc, and Ti. In this study, Mn was selected as the "conservative element" (Bergamaschi et al., 2002; Salati and Moore, 2009; Duodu et al., 2016; Islam et al., 2018). Before adopting EF, PCA and CA were used to determine whether Mn is a conserved element (Han et al., 2006). This approach can negate the effect of terrestrial sedimentary input and identify anthropogenic pollution sources (Brady et al., 2015).

This was determined using the following equation given by Muller (1981):

$$EF = \frac{(C_i/C_b)_{\text{sample}}}{(C_i/C_b)_{\text{background}}}, \tag{2}$$

where C_i is the concentration of the element of interest and C_b is the concentration of the normalization element. Five contamination categories are associated with EF by Sutherland (2000): EF < 2 indicates depletion to minimum enrichment; 2 ≤ EF < 5 represents moderate enrichment; 5 ≤ EF < 20 indicates significant enrichment; 20 ≤ EF < 40 represents very high enrichment; and EF ≥ 40 indicates extremely high enrichment.

2.4.2 Multi-Element Indices

The limitations of the single element indices have led to the development of multiple element indices to assess sediment quality (Brady et al., 2015; Duodu et al., 2016). The most common multiple element indices are the modified contamination index (mCd) developed by Hakanson (1980), the Nemerow pollution index (PI) (Nemerow, 1991), and the modified pollution index (MPI) by Brady et al. (2015).

The mCd calculates the average impact of all heavy metal elements at a sampling point to evaluate sediment pollution. Meanwhile, the PI allows the qualification of sediment quality that is much more considerate of the effect of a single element by using a weighted average. To account for the behaviour of sediments within estuaries and the possibilities of multiple sediment sources, an improved method for determining the PI is proposed by using EF to calculate an MPI, which would allow for the non-conservative behaviour of sediments due to normalization against an element (Brady et al., 2015). **Eqs. 4** and **6** show how mCd, PI, and the MPI are calculated. The thresholds for sediment quality classification using the three integrated indices are presented in **Table 1**.

$$\mathrm{mC_d} = \frac{\sum_{i=1}^{n} CF_i}{\mathrm{n}}, \tag{3}$$

$$PI = \sqrt{\frac{\left(\mathrm{CF_{average}}\right)^2 + \left(\mathrm{CF_{max}}\right)^2}{2}}, \tag{4}$$

$$MPI = \sqrt{\frac{\left(\mathrm{EF_{average}}\right)^2 + \left(\mathrm{EF_{max}}\right)^2}{2}}, \tag{5}$$

where CF_i, $CF_{average}$, CF_{max}, and EF_{max} represent contamination factors for an individual element, average of CF, average of EF, maximum CF, and maximum EF, respectively.

2.4.3 Ecological Risk Assessment

The SQGs is a simple and comparative indicator used to assess the quality of sediments and their adverse biological effects on the aquatic ecosystem (MacDonald et al., 2000; Ke et al., 2017; Kang et al., 2020). In this study, two limit values were applied to evaluate the potential risk of the ecosystem, threshold effect content (TEC) and probable effect content (PEC). If the content is below the TEC, then adverse biological effects rarely occur; if the concentrations are equal to or greater than the TEC but less than the PEC, then a range of biological effects occasionally occur; if the concentrations are at or above the PEC, then a probable effect range of adverse biological effects frequently occur (Ke et al., 2017; Kang et al., 2020).

Potential ecological risk index (RI) and modified potential ecological risk index (MRI) were used to assess the extent of heavy metal pollution and its potential ecological harm (Hakanson, 1980). RI takes into consideration the CF of metal, potential ecological risk factors (E_r), and toxicological response factors (T_r) (1 for Zn; 2 for Cr; 5 for Cu, Pb, and Ni; and 30 for Cd) (Hakanson 1980; Xu et al., 2008). MRI uses EF in the calculation of RI to account for the effect of terrestrial sedimentary (Duodu et al., 2016). **Eqs. 7** and **8** show how the RI and MRI are calculated.

$$RI = \sum_{i=1}^{n} E_r^i = \sum_{i=1}^{n} T_r^i \times CF_i, \tag{6}$$

$$MRI = \sum_{i=1}^{n} E_r^i = \sum_{i=1}^{n} T_r^i \times EF_i, \tag{7}$$

TABLE 1 | Thresholds for sediment quality classification for multi-element indices.

Class	Sediment qualification	mCd	PI	MPI
0	Unpolluted	MCd < 1.5	PI < 0.7	MPI < 1
1	Slightly polluted	1.5 ≤ mCd < 2	0.7 ≤ PI < 1	1 ≤ MPI < 2
2	Moderately polluted	2 ≤ mCd < 4	1 ≤ PI <2	2 ≤ MPI < 3
3	Moderately–heavily polluted	4 ≤ mCd < 8	–	3 ≤ MPI < 5
4	Severely polluted	8 ≤ mCd < 16	2 ≤ PI < 3	5 ≤ MPI < 10
5	Heavily polluted	16 ≤ mCd < 32	PI ≥ 3	MPI ≥ 10
6	Extremely polluted	MCd ≥ 32	–	–

TABLE 2 | The heavy metal consensus-based quality guideline values (mg/kg) and background values (mg/kg) in soils of Jian and Moyang river basins.

	Zn	Pb	Ni	Mn	Cu	Cr	Cd
Mean	58.43	40.28	26.57	144.97	24.21	6.71	0.17
Range	18.64–109.34	21.16–97.80	9.89–53.66	57.86–317.41	1.20–93.66	1.27–11.44	0.004–0.63
VC	0.38	0.38	0.42	0.48	0.85	0.36	0.93
Skewness	0.52	1.55	0.57	0.77	1.60	–0.66	1.17
Background (Shi et al., 2016)	67.60	28.49	20.13	575.35	19.00	51.30	0.13
Multiple	0.86	1.41	1.32	0.25	1.27	0.13	1.30

TABLE 3 | Metal concentrations (mg/kg) in the Jian and Moyang rivers compared with the mean sediment concentrations reported previously.

Location	Zn	Pb	Ni	Mn	Cu	Cr	Cd	References
The Jian River and Moyang River, China	109.34	97.80	53.66	317.41	93.66	11.44	0.63	Present study
The Pearl River, China	543.60	104.58	54.10	1104.73	80.20	86.62	10.60	Xie et al. (2012)
The Yangtze River, China	98.09	27.72	–	–	40.47	74.60	0.30	Ke et al. (2017)
The Haihe River, China	89.41	20.65	44.63	–	74.23	92.09	0.264	Kang et al. (2020)
Langat River, Malaysia	29.71	15.52	–	–	79.10	–	0.10	Shafie et al. (2014)

TABLE 4 | Loading corresponding to three factors for surface sediments of the Jian and Moyang rivers.

Parameters	Component		
	Factor 1	Factor 2	Factor 3
Cr	0.897		
Cu	0.845		
Ni	0.724		
Zn	0.723		
Pb	0.532		
Mn		0.903	
Fe		0.845	
Cd			0.889
Eigenvalue	4.424	3.794	2.916
Total variance (%)	37.872	29.348	19.165

where E_r^i is the potential ecological risk index of an individual element, T_r^i is the biological toxic response factor of an individual element, CF_i is the contamination factor for each single element, and EF_i is the enrichment factor for each single element. Grading of potential ecological risk factors are as follows: E_r <40 for low risk, 40–80 indicates moderate risk, 80–160 represents considerable risk, 160–320 represents high risk, and >320 represents very high risk. Grades of potential and modified ecological risk index such as RI or MRI <150 indicates low risk, 150–300 represents moderate risk, 300–600 indicates considerable risk, and >600 indicates very high risk.

3 RESULTS

3.1 Concentration of Heavy Metals in Surface Sediments

Table 2 illustrates the descriptive statistics of the heavy metal contents of the Jian and Moyang rivers. The metals in the sediments from the Jian and Moyang rivers ranged from 57.86 to 317.41 for Mn, 18.64 to 109.34 for Zn, 21.16 to 97.80 for Pb, 9.89 to 53.66 for Ni, 1.20 to 93.66 for Cu, 1.27 to 11.44 for Cr, and 0.004–0.63 mg/kg for Cd. The values of the median concentrations can be ranked in descending order for the Jian and Moyang rivers: Mn > Zn > Pb > Ni > Cu > Cr > Cd. The comparison of the reported results of this study with the background value of water system sediments in the South China Youjiang orogenic zone and low hill water system sediments (Shi et al., 2016) revealed that the concentrations of Cu, Cd, Ni, and Pb were higher than the background value. The average Zn, Pb, Cu, and Cd concentrations in this study were higher than the other studies in **Table 3**, whilst the concentrations of Zn, Pb, and Cd were much lower in China's Pearl River.

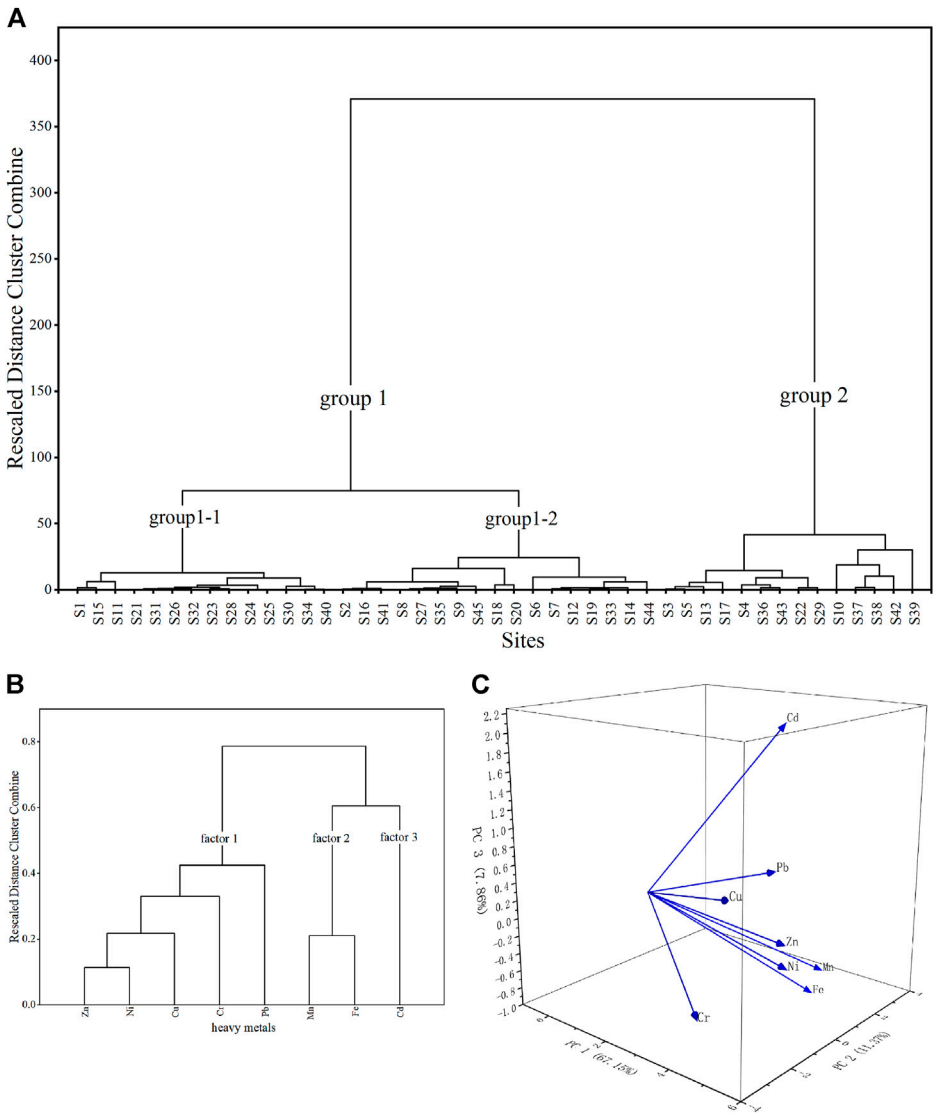


FIGURE 2 | (A–B) Cluster diagram of sampling sites and heavy sites and heavy metal elements obtained using Ward's method and Squared Euclidean distance matrix. **(A)** Sampling sites. **(B)** Heavy metal elements. **(C)** Principal component analysis of different metals.

TABLE 5 | Evaluation results in contamination factor (CF) and enrichment factor (EF) percentage (%) in the surface sediments of the Jian and Moyang rivers.

Contamination factor (CF)							
Ranges of indexes	**State of pollution**	**Zn**	**Pb**	**Ni**	**Cu**	**Cr**	**Cd**
Average		0.86	1.41	1.32	1.27	0.13	1.30
CF < 1	Low contamination	62.22%	20.00%	31.11%	44.44%	100.00%	53.33%
1 ≤ CF < 3	Moderate contamination	37.78%	77.78%	68.89%	46.67%	0.00%	33.33%
3 ≤ CF < 6	Considerable contamination	0.00%	2.22%	0.00%	8.89%	0.00%	13.33%
CF ≥ 6	Very high contamination	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Enrichment factor (EF)							
	Average	3.83	6.40	5.72	5.39	0.61	5.47
EF < 2	Depletion to minimum enrichment	4.44%	0.00%	0.00%	22.22%	0.00%	24.44%
2 ≤ EF < 5	Moderate enrichment	82.22%	35.56%	44.44%	31.11%	0.00%	40.00%
5 ≤ EF < 20	Significant enrichment	13.33%	62.22%	55.56%	44.44%	0.00%	31.11%
20 ≤ EF < 40	Very high enrichment	0.00%	2.22%	0.00%	2.22%	0.00%	4.44%
EF > 40	Extremely high enrichment	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

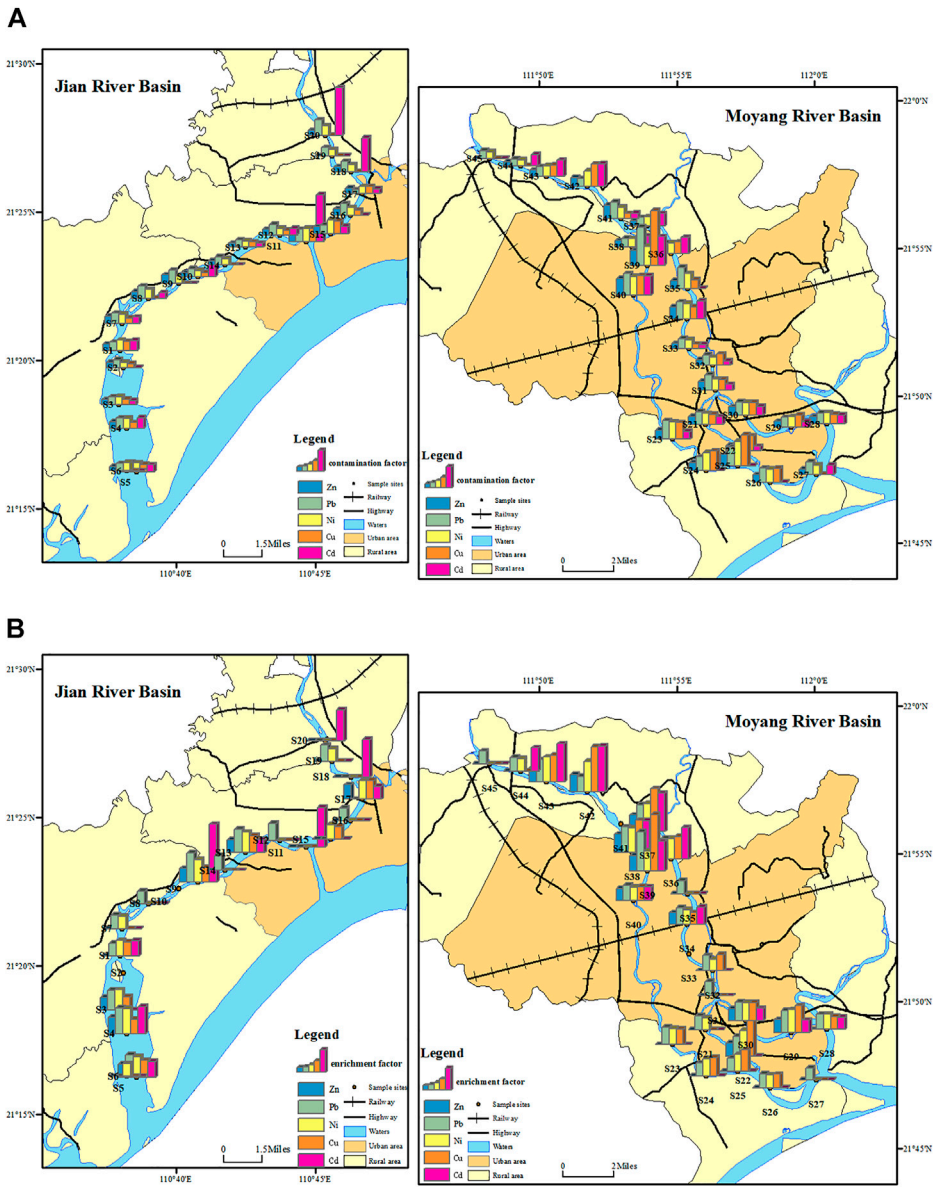


FIGURE 3 | Distribution of pollution evaluation results by CF and EF from Jian and Moyang rivers' sediments. **(A)** Contamination factor; **(B)** enrichment factor (The results of the contamination factor only show the sampling sites with moderate contamination and above, and the results of the enrichment factor only show the sampling sites with moderate contamination and above).

3.2 Cluster Analysis and Principal Component Analysis

Cluster analysis (CA) was performed to divide the sampling sites into groups with similar concentrations of heavy metals in sediments. CA is a statistical method used to identify groups or clusters of similar parameters based on similarities within a class and dissimilarities between different classes. In this study, CA made it possible to divide the measurement points into groups 1–1, 1–2, and 2 (**Figure2**).

Considering the relatively high variability of the heavy metal concentrations in certain groups of sampling sites, principal component analysis (PCA) and CA was separately performed for three clusters (**Table 4** and **Figure 2**). The first factor represented 37.872% out of the total variance and members of the significant variables with high loading on Cr, Cu, Ni, Zn, and Pb. The second factor was dominated by Mn and Fe, accounting for 29.348% of the total variance. The third factor contributed 19.165% of the total variance and was only characterized by the loading of Cd.

3.3 Single Pollution Indices

Amongst the studied metals, the average values of CF decreased in the order Pb > Ni > Cd > Cu > Zn > Cr. Meanwhile, Zn, Pb Ni,

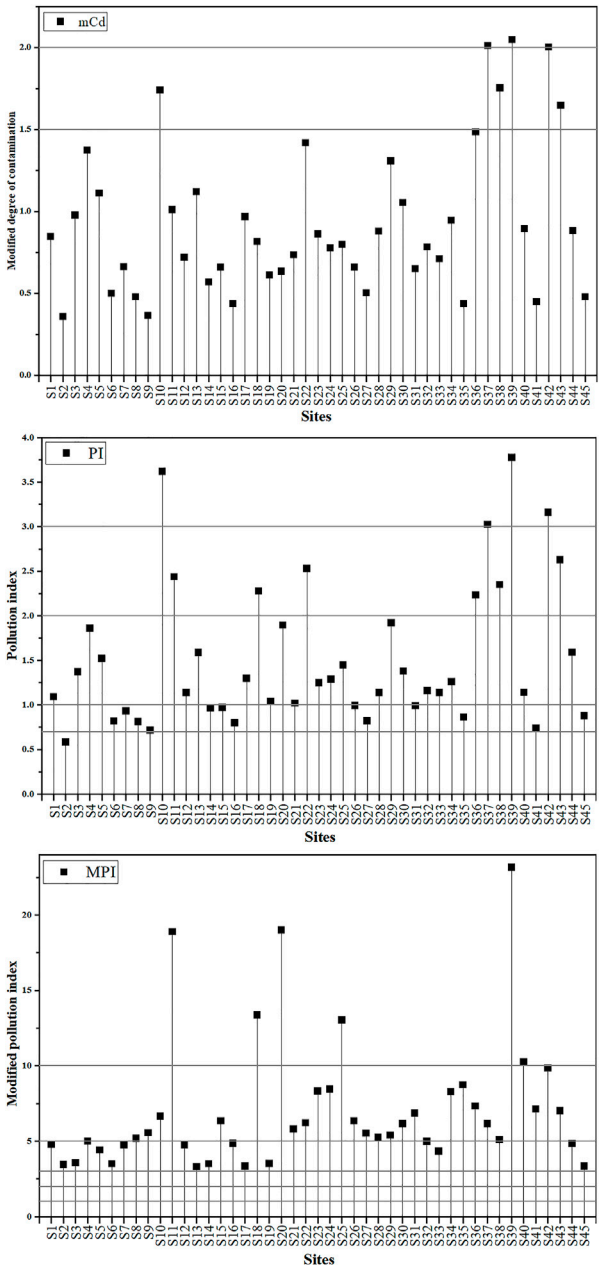


FIGURE 4 | Sediment quality assessment by multi-element indices (mCd: modified degree of contamination, PI: pollution index, and MPI: modified pollution index).

Cu, and Cd showed moderate contamination in some sample stations (37.78%, 77.78%, 68.89%, 46.67.56, and 33.33%, respectively) (**Table 5**). Significant contamination was observed at the S22, S17, S19, and S22 (group2) sample stations, mainly because of Cu concentration. At sample stations S11 (group 1–1), S18 (group 1–2), and S10, S37, S42, and S43 (group 2), the values of CF for Cd indicated considerate contamination. In addition, S39 (Group2) sample station indicated considerable contamination by Pb (**Figure 3**).

Group 2 consisted of Mn–Fe in PCA, which widely exists in nature (Graney and Eriksen, 2004). This notion means that Mn is a conservative element in the studied environment. EF values computed for the metals at various sites are also presented in **Table 5**. The average values of EF decreased in the order Pb > Ni > Cd > Cu > Zn > Cr, which is similar to the CF results. The assessment of the EF values for the Jian and Moyang rivers demonstrated that Pb, Ni, Cd, and Cu have greater enrichment compared with the other heavy metals (**Table 5**). In the case of the Jian and Moyang rivers, the values of EF for Pb indicated moderate enrichment at 35.56% of the sample stations and significant enrichment at 62.22% of the sample stations, whilst very high enrichment was observed at S39 (group 2). The analysis of the EF for Ni shows that moderate enrichment occurs at 44.44% of the sample stations and significant enrichment at 55.56% of the sample stations. The assessment of the EF for Cd shows moderate enrichment at 40.00% of the sample stations and significant enrichment at 31.11% of the sample stations, but very high enrichment at S11 (group 1–1) and S20 (group 1–2). In addition, 31.11% of the sample stations indicated moderate enrichment, and 44.44% of the sample stations implied significant enrichment by Cu, with only S39 (group2) having very high enrichment (**Figure 3**). Spatial distribution of the CF and EF from Jian and Moyang rivers show heavier metal enrichment of sediments in the upper course of Moyang River and heavier Cd metal enrichment of sediments in the upper course of Jian River (**Figure 3**).

3.4 Multi-Element Indices

Given that the problem of heavy metal pollution in sediments is influenced by various heavy metals, this study used multi-element indices (mCd, PI, and MPI) to assess the influence of multiple contamination species at a site to overcome some limitations of the single indices. In this study, contamination was low, as indicated by "unpolluted" (**Figure 4**), because none of the sites had an index (mCd) greater than 1.5 except S10, S17, S18, S23, S22, and S39 (group 2), which are slightly to moderately polluted (**Table 6**). However, the PI value for the Jian River (group 1–2:S2) was lower than 0.7 at only one sample station (**Figure 4**). The PI values for the rest of the sample stations were above 0.7, indicating that the overall contamination status of the Jian and Moyang rivers was polluted. In the case of the study area, 28.89 and 46.67% of the sample stations were slightly polluted and moderately polluted by PI, respectively. The analysis of the PI for the Jian and Moyang rivers shows that severe pollution occurs at 13.33% of the sample stations (group 1–1: S11, group 1–2: S18, group2: V2, V16, V18, and V23), whilst heavy pollution was observed only at four sample stations (group 2: S10, S37, S39, and S42) (**Figure 4**). The PI results indicated heavier pollution of sediments in group 2 sample stations. The values of MPI for sediments indicated that the pollution was moderate to heavily in the study area, with a mean value of 7.02. The MPI values for 35.56% of the sample stations indicate moderate to heavy pollution, and 51.11% denoted heavy pollution (**Table 6**). Moreover, assessment of the MPI shows that extreme pollution occurs at 13.33% of the sample stations (group1–1: S11 and S25; group1–2: S18 and S20; and group2: S39 and S40).

TABLE 6 | Variations of evaluation results in the modified degree of contamination (mCd), pollution index (PI), and modified pollution index (MPI) percentage (%) in the surface sediments of the Jian and Moyang rivers.

State of pollution	Ranges of indexes	mCd	Ranges of indexes	PI	Ranges of indexs	MPI
	Average	0.94	Average	1.52	Average	7.02
Unpolluted	MCd < 1.5	86.67%	PI < 0.7	2.22%	MPI < 1	0.00%
Slightly polluted	1.5 < mCd < 2	6.67%	0.7 < PI < 1	28.89%	1 < MPI < 2	0.00%
Moderately polluted	2 < mCd < 4	6.67%	1 < PI < 2	46.67%	2 < MPI < 3	0.00%
Moderately-heavily polluted	4 < mCd < 8	0.00%			3 < MPI < 5	35.56%
Severely polluted	8 < mCd < 16	0.00%	2 < PI < 3	13.33%		
Heavily polluted	16 < mCd < 32	0.00%	PI > 3	8.89%	5 < MPI < 10	51.11%
Extremely polluted	MCd > 32	0.00%			MPI > 10	13.33%

3.5 Ecological Risk Assessment of Metals in Sediment

When compared with the reference limit values of TEC and PEC, the majority of the samples were below the PEC value, which is defined as the limit above which a toxic effect on aquatic can be expected. Only 4.44% of the Ni sediment samples for the Jian and Moyang rivers were above this level (**Table 7**). However, biological Ni, Cu, and Ni effects occur in 51.11%, 31.11%, and 4.44% of the sample, with values between TEC and PEC. Therefore, the application of SQGs to the sediment analyzed revealed that Pb, Ni, and Cu can potentially cause biological effects.

The E_r calculated using CF showed that the heavy metals all have a low risk, except for Cd (**Table 8**). In the case of the study rivers, the E_r with CF for Cd indicated moderate to considerate risk at 28.89% of the sample stations. Meanwhile, the average value of the E_r with EF showed that only Cd had an ecological risk in the study area. The obtained results showed that the E_r with EF for Pb, Ni, and Cu has moderate to considerable ecological risk in some sampling stations (20.00%, 15.56%, and 13.33%, respectively) (**Table 8**).

The RI and MRI were computed (**Figure 5**). The RI values for the analyzed heavy metals were lower than 150, indicating that none of the sample stations has an ecological risk for the aquatic environment (**Table 9**), except for S10 and V22 (Group2). However, the mean values of the MRI indicated that moderate ecological risk occurred in Jian and Moyang rivers. In the case of the study area, 48.89% of the sample stations showed moderate risk. Moreover, the MPI values were more than 300 at 22.22% of the sample stations (group 1–1: S11 and S34; group1–2: S18 and S20; and group2: S10, S36, S39, S40, S42, and S43).

4 DISCUSSION

4.1 Contamination and Risk Assessment

In this study, the mean values of the heavy metals in CF and EF showed that Pb, Ni, Cd, and Cu are moderate contaminants. This finding is similar to the results of related studies in the Pearl River Basin (Xiao et al., 2013; Zhao et al., 2016; Jiao et al., 2018). The Pb contamination can be linked to leaded gasoline (Al-Khashman, 2007) because the seriously impacted sample station S39 is close to the highway. The proximity of S10, S11, S18, S37, S42, and S43 to the farming area indicates that Cd contamination is a result of the application of pesticides and fertilizers (Gray et al., 1999). The considerate Cu contamination may have been caused by antifouling paints from a shipping yard and ferry activities near S22 (Duodu et al., 2016) and the use of Cu-containing insecticides and fungicides around S22, S37, and S39. Moreover, Cu contamination may be associated with traffic, such as tyre and brake wear because the affected sample station S39 was a road (Duodu et al., 2016). According to the SQGs, some sample stations may occasionally have biological Ni, Pb, and Cd effects. This result is similar to the research results in the Pearl River Basin (Niu et al., 2009; Xiao et al., 2013; Liao et al., 2017; Yao et al., 2021). The results of the potential ecological risk indices of each metal (Er) showed that all heavy metals are a low risk, except for Cd, based on the mean value in Er with (EF). This difference is probably related to the magnitude of toxicological response factors:1 for Zn; 2 for Cr; 5 for Cu, Pb, and Ni; and 30 for Cd (Hakanson 1980; Xu et al., 2008). Cd is highly toxic to living organisms (Kumar et al., 2018). Our study performed a PCA to extract a small number of latent factors to analyze the relationships amongst variables. The results of the PCA, in combination with the abovementioned research, on the sources of single heavy metals in the study are as follows. The first factor is dominated by Cr, Cu, Ni, Zn, and Pb, which are associated with a variety of polluted sources, such as transportation, agricultural, industrial, and natural sources. The second factor showed strong loadings of Mn and Fe, indicating a lithogenic contribution (Graney and Eriken, 2004). The third factor is characterized by Cd, demonstrating the impact of agricultural production.

To identify the groups of similar sample stations, CA was used to divide the sample sites in this study into groups 1–1, 1–2, and 2. In multi-element indices, S11, S23, and S25 (group 1–1), S18 and S20 (group 1–2), and S10, S17, S22, S36, S37, S38, S39, S40, S42, and S43 (group 2) are slightly to moderately polluted by mCd or severely polluted by PI or extremely polluted by MPI. This notion indicates that group 2 is the most prominent in heavy metal pollution. The analysis of the RI showed that moderate ecological risk occurs at S10 and S42 (group 2). The result of the MRI indicated that very high ecological risk occurs at S11 and S34 (group 1–1), S18 and S20 (group 1–2), and S10, S16, S39, S40, S42, and S43 (group 2). The comprehensive pollution of group 2 is more serious than that of group 1, and 65.16% of the sampling

TABLE 7 | Comparison between sediment quality guidelines (SQGs) and heavy metal concentrations (mg/kg) of all samples in Jian and Moyang rivers.

	Zn	Pb	Ni	Cu	Cr	Cd
% of samples < TEC	100	51.11	44.44	80.00	100	100
% of samples between TEC–PEC		48.89	51.11	31.11		
% of samples >PEC			4.44			

points with significant and very high Cd enrichment are located in group 2. This finding indicated that group 2 might be affected by agricultural production mainly located in the upper reaches of Moyang River. Although groups 1–1 and 1–2 have no significant difference in the proportion of high pollution values, the MRI evaluation results showed that group 1–1 was moderate and above an ecological pollution level, except S26, and 52.94% of the samples in group 1–2 are low risk. In the single element Indices, all the samples in group 1–1 have Cu and Ni significant enrichment, except S1 and S32. In combination with the sample distribution, group 1–1 is mainly affected by industrial production and urban activities. These findings suggest that continuous monitoring of Cd, Pb, Ni, and Cu in the river sediment of the Jian and Moyang rivers should be directed to evaluate the threat of these elements to the public health and to the ecology of the river under study. Agricultural pollution must be given attention, and the detection of industrial and urban pollution in river governance must be strengthened.

4.2 Use of Pollution Indices

The RI evaluates the combined pollution risk of an aquatic system through a toxic-response factor using CF. Although the CF, PI, and RI have been increasingly used in recent years, there are some limitations relating to their use. CF does not take into consideration lithogenic and sedimentary inputs of the element of interest. If the CF is used to determine the PI and RI, this could be a source of error when estuarine environments are considered because there are areas of intense sedimentation with significant input from terrestrial waterways (Brady et al., 2015; Duodu et al., 2016). To negate the effect of terrestrial sedimentary input, EF can be used to standardize the impact of terrestrial inputs by normalizing the element of interest against an element that has no anthropogenic source. An improved method for determining PI and the RI is proposed by using EF to calculate MPI and MRI, allowing for the non-conservative behaviour of sediments due to normalization against an element (Brady et al., 2015).

In this study, six heavy metal contaminants were used to calculate the CF, EF, mCd, PI, MPI, RI, and MRI. Although the CF and EF indicated heavy metal contamination in this study, EF appeared to detect contamination (at more sites and of more metals) better than CF in the worst-case scenario. The assessment of the CF values demonstrated that Cd was considered contamination in S11. Meanwhile, this sample station has extremely high enrichment based on the EF. This finding suggests that the use of EF is more sensitive to monitoring heavy metal pollution because it can normalize the impact of terrestrial sedimentary inputs to provide more useful information. Similar results were obtained in earlier studies (Duodu et al., 2016; Kumar et al., 2018), wherein EF could detect terrestrial sedimentary inputs of metals due to the normalization. In multi-element indices, the mCd values for 86.67% of the sample stations indicate unpolluted, whilst PI overestimates the risk at all sites and indicates that only 2.22% of the sample stations are unpolluted. The weighted average rather than just the average CF of PI makes a more likely, higher value for the index, and the trigger thresholds are lower than those for the other indices (Brady et al., 2015). Accordingly, the PI is more likely to identify high levels of contamination for a suite of elements. The results of MPI also indicated that the sample stations were affected by multiple contaminations. However, MPI shows more sample points with high values. According to the ecological risk assessment of the metals in the sediment, the RI showed no ecological risk for the aquatic environment, except for S10 and V22. However, the MRI values indicated that 71.11% of the sample stations were at moderate to very high ecological risk. Therefore, the MPI and MRI, which are calculated from EF, can more sensitively detect pollution in this study than PI and MRI. The results of the study would be valuable for researchers in environmental quality evaluation, and the applied methods applied can be used for pollution assessment in other environments.

4.3 Pollution Assessment in Small Rivers

Environmental conditions are complicated, with towns, villages, and farmland on both sides of small rivers. Small rivers have been

TABLE 8 | Evaluation results in the potential ecological risk index of an individual element (Eri) with EF or CF percentage (%) in the surface sediments of the Jian and Moyang rivers.

Ranges of indexes	State of pollution	Zn (CF)	Zn (EF)	Pb (CF)	Pb (EF)	Ni (CF)	Ni (EF)	Cu (CF)	Cu (EF)	Cr (CF)	Cr (EF)	Cd (CF)	Cd (EF)
Average		0.86	3.83	7.07	32.02	6.60	28.59	6.37	26.96	0.26	1.23	38.85	164.10
<40	Low risk	100.00%	100.00%	100.00%	80.00%	100.00%	84.44%	100.00%	86.67%	100.00%	100.00%	68.89%	20.00%
40–80	Moderate risk	0.00%	0.00%	0.00%	17.78%	0.00%	15.56%	0.00%	8.89%	0.00%	0.00%	11.11%	15.56%
80–160	Considerable risk	0.00%	0.00%	0.00%	2.22%	0.00%	0.00%	0.00%	4.44%	0.00%	0.00%	17.78%	31.11%
160–320	High risk	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	22.22%
>320	Very high risk	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	11.11%

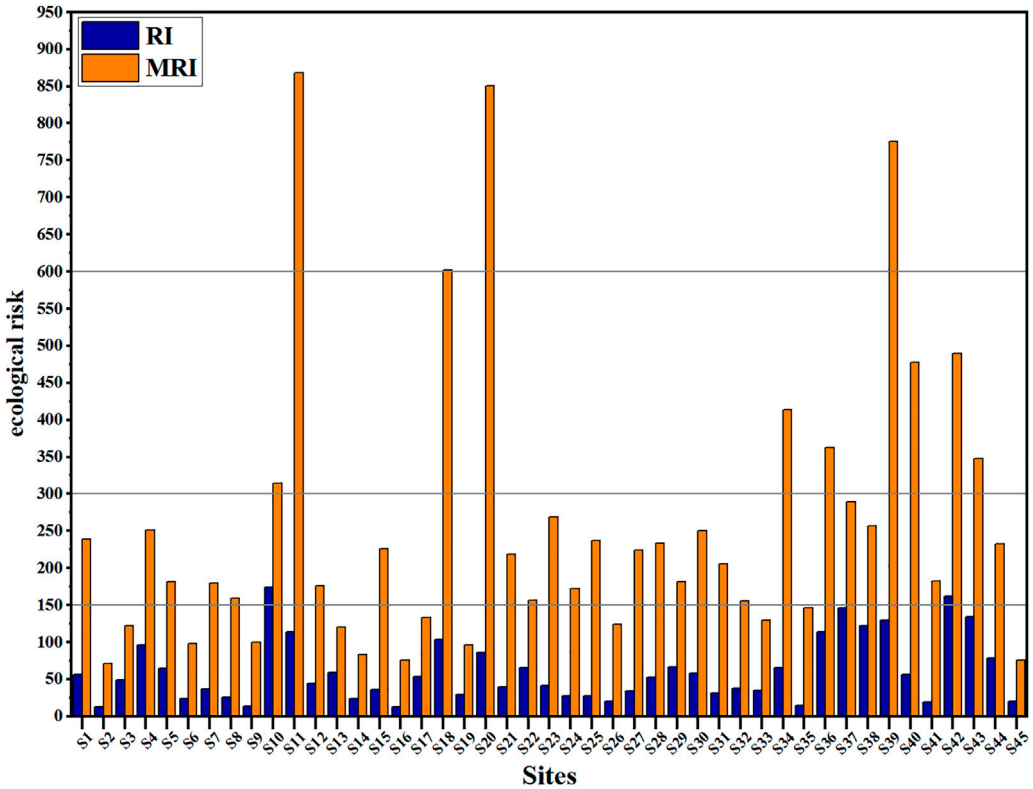


FIGURE 5 | Sediment quality assessment by ecological risk assessment of metals in sediment. (RI: potential ecological risk index and MRI: modified ecological risk index).

TABLE 9 | Variations of evaluation results in the potential ecological risk index (RI) and modified ecological risk index (MRI) percentage (%) in the surface sediments of the Jian and Moyang rivers.

Ranges of indexes	State of pollution	RI	MRI
Average		60.27	256.72
<150	Low risk	95.56%	28.89%
150–300	Moderate risk	4.44%	48.89%
300–600	Considerable risk	0.00%	13.33%
>600	Very high risk	0.00%	8.89%

used as a drainage channel for intensive runoff in densely populated urban areas or as a water source in agricultural areas (Lee et al., 2022). Once polluted, small rivers are difficult to recover because they are low in volume, slow in flow, and heavily silted. If small rivers are polluted, then the pollutants flow into large rivers and oceans, affecting ecological security. In comparison with big rivers, small rivers are more sensitive to changes in the surrounding environment and capable of tending to human activities and natural factors (Zhou et al., 2020). Therefore, precise metal source identification, distribution, and pollution characterization are particularly important in small rivers.

The findings of Nasrabadi et al. (2010), Giri et al. (2013), Li et al. (2014), Wu et al. (2017), and Arisekar et al. (2022) on the small-scale rivers indicate the impact of anthropogenic agents on the heavy metal abundance in Xiaoyang, Thamirabarani, Subarnarekha, Lianshan, Wuli, and Cishan river sediments. Wu et al. (2017) showed that anthropogenic input is associated with urban development, especially industrial intensification. Giri et al. (2013) confirmed that the increased concentration of metals in the Subarnarekha River is due to the direct discharge of industrial, urban, and mining wastes into the river. However, Arisekar et al. (2022) highlighted that the high level of pollution was influenced by agricultural runoff. According to this study, the Moyang and Jian rivers are moderately to very highly contaminated with Pb, Ni, Cd, and Cu metals when the CF, EF, mCd, PI, MPI, RI, and MRI values are taken into consideration. When combined with PA and PCA, we concluded that the increased concentrations of metals in the sediment of the Moyang and Jian rivers are due to the discharge of agricultural, industrial, and urban wastes into the river. Therefore, small rivers can be susceptible to the influence of human activities, which must be addressed.

5 CONCLUSION

The concentrations of metals in the Jian and Moyang rivers' sediment are presented in this study. Two single contamination indices (EF and CF) were used to assess the

sediment quality in addition to the three multi-element indices (mCd, PI, and MPI). The river sediment was slightly to extremely polluted. The SQGs revealed that Pb, Ni, and Cu have potential biological effects. MPI, which took into consideration the lithogenic and sedimentary inputs of the element of interest, indicated that most sampling points have moderate to very high ecological risks. CA and PCA showed that pollution in this study area is primarily caused by anthropogenic activities (agricultural pollution, industrial effluents, and sewage discharge). Our study found that both sides of small rivers are vulnerable to being polluted by production activities, which must be taken seriously. In comparison with the various evaluation methods, the indices calculated from EF detect pollution more sensitively and may also be used to assess pollution in other environments.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**; further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

JF: conception, experimentation, analysis of data, and writing—original draft; YL: pretreatment of samples and experimentation; ML: collection of samples, experimentation, and data analysis; TO: conception, methodology, and writing—review and editing; MY: collection of samples, conception, data curation, methodology, and writing—review and editing.

FUNDING

This research was funded by the Natural Science Foundation of Guangdong Province (Grant Nos. 2018B030311056 and 2019A1515110195).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fenvs.2022.927765/full#supplementary-material

REFERENCES

Ahmed, F., Fakhruddin, A. N. M., Imam, M. D. T., Khan, N., Khan, T. A., Rahman, M. M., et al. (2016). Spatial Distribution and Source Identification of Heavy Metal Pollution in Roadside Surface Soil: a Study of Dhaka Aricha Highway, Bangladesh. *Ecol. Process* 5 (1). doi:10.1186/s13717-016-0045-5

Akcay, H., Oguz, A., and Karapire, C. (2003). Study of Heavy Metal Pollution and Speciation in Buyak Menderes and Gediz River Sediments. *Water Res.* 37 (4), 813–822. doi:10.1016/s0043-1354(02)00392-5

Al-Khashman, O. A. (2007). Determination of Metal Accumulation in Deposited Street Dusts in Amman, Jordan. *Environ. Geochem Health* 29 (1), 1–10. doi:10.1007/s10653-006-9067-8

Arisekar, U., Shakila, R. J., Shalini, R., Jeyasekaran, G., Keerthana, M., Arumugam, N., et al. (2022). Distribution and Ecological Risk Assessment of Heavy Metals Using Geochemical Normalization Factors in the Aquatic Sediments. *Chemosphere* 294, 133708. doi:10.1016/j.chemosphere.2022.133708

Bergamaschi, L., Rizzio, E., Valcuvia, M. G., Verza, G., Profumo, A., and Gallorini, M. (2002). Determination of Trace Elements and Evaluation of Their Enrichment Factors in Himalayan Lichens. *Environ. Pollut.* 120 (1), 137–144. doi:10.1016/s0269-7491(02)00138-0

Brady, J. P., Ayoko, G. A., Martens, W. N., and Goonetilleke, A. (2015). Development of a Hybrid Pollution Index for Heavy Metals in Marine and Estuarine Sediments. *Environ. Monit. Assess.* 187 (5). doi:10.1007/s10661-015-4563-x

Budianta, W. (2020). Heavy Metal Pollution and Mobility of Sediment in Tajum River Caused by Artisanal Gold Mining in Banyumas, Central Java, Indonesia. *Environ. Sci. Pollut. Res.* 28 (7), 8585–8593. doi:10.1007/s11356-020-11157-5

Chen, W. T. (2014). Analysis on the Status Quo and Water Quality Change Trend of Water Functional Zones in the Moyang River Basin. *Guangdong Water Resour. Hydropower* (10), 31–34. doi:10.11905/j.issn.1008-0112.2014.10.009

Davutluoglu, O. I., Seckin, G., Ersu, C. B., Yilmaz, T., and Sari, B. (2011). Heavy Metal Content and Distribution in Surface Sediments of the Seyhan River, Turkey. *J. Environ. Manag.* 92 (9), 2250–2259. doi:10.1016/j.jenvman.2011.04.013

Duodu, G. O., Goonetilleke, A., and Ayoko, G. A. (2016). Comparison of Pollution Indices for the Assessment of Heavy Metal in Brisbane River Sediment. *Environ. Pollut.* 219, 1077–1091. doi:10.1016/j.envpol.2016.09.008

Fu, J., Hu, X., Tao, X., Yu, H., and Zhang, X. (2013). Risk and Toxicity Assessments of Heavy Metals in Sediments and Fishes from the Yangtze River and Taihu Lake, China. *Chemosphere* 93 (9), 1887–1895. doi:10.1016/j.chemosphere.2013.06.061

Gaiero, D. M., Ross, G. R., Depetris, P. J., and Kempe, S. (1997). Spatial and Temporal Variability of Total Non-residual Heavy Metals Content in Stream Sediments from the Suquia River System, Cordoba, Argentina. *Water Air Soil Pollut.* 93 (1-4), 303–319. doi:10.1007/bf02404763

Giri, S., Singh, A. K., and Tewary, B. K. (2013). Source and Distribution of Metals in Bed Sediments of Subarnarekha River, India. *Environ. Earth Sci.* 70 (7), 3381–3392. doi:10.1007/s12665-013-2404-1

Graney, J. R., and Eriksen, T. M. (2004). Metals in Pond Sediments as Archives of Anthropogenic Activities: a Study in Response to Health Concerns. *Appl. Geochem.* 19 (7), 1177–1188. doi:10.1016/j.apgeochem.2004.01.014

Gray, C. W., McLaren, R. G., Roberts, A. H. C., and Condron, L. M. (1999). The Effect of Long-Term Phosphatic Fertiliser Applications on the Amounts and Forms of Cadmium in Soils under Pasture in New Zealand. *Nutrient Cycl. Agroecosyst.* 54 (3), 267–277. doi:10.1023/A:1009883010490

Hakanson, L. (1980). An Ecological Risk Index for Aquatic Pollution control.a Sedimentological Approach. *Water Res.* 14 (8), 975–1001. doi:10.1016/0043-1354(80)90143-8

He, M., Wang, Z., and Tang, H. (1997). Spatial and Temporal Patterns of Acidity and Heavy Metals in Predicting the Potential for Ecological Impact on the Le an River Polluted by Acid Mine Drainage. *Sci. Total Environ.* 206 (1), 67–77. doi:10.1016/s0048-9697(97)00217-9

Hoang, H.-G., Lin, C., Chiang, C.-F., Bui, X.-T., Lukkhasorn, W., Bui, T.-P. -T., et al. (2021). The Individual and Synergistic Indexes for Assessments of Heavy Metal Contamination in Global Rivers and Risk: a Review. *Curr. Pollut. Rep.* 7, 247–262. doi:10.1007/s40726-021-00196-2

Islam, M. S., Hossain, M. B., Matin, A., and Islam Sarker, M. S. (2018). Assessment of Heavy Metal Pollution, Distribution and Source Apportionment in the Sediment from Feni River Estuary, Bangladesh. *Chemosphere* 202, 25–32. doi:10.1016/j.chemosphere.2018.03.077

Jiao, Z., Li, H., Song, M., and Wang, L. (2018). Ecological Risk Assessment of Heavy Metals in Water and Sediment of the Pearl River Estuary, China. *IOP Conf. Ser. Mat. Sci. Eng.* 394, 052055. doi:10.1088/1757-899x/394/5/052055

Kang, D., Zheng, G., Yu, J., Chen, Q., Zheng, X., Zhong, J., et al. (2021). Hydropower Reservoirs Enhanced the Accumulation of Heavy Metals towards Surface Sediments and Aggravated Ecological Risks in Jiulong River Basin, China. *J. Soils Sediments* 21 (10), 3479–3492. doi:10.1007/s11368-021-03002-0

Kang, M., Tian, Y., Zhang, H., and Lan, Q. (2020). Distribution, Ecological Risk Assessment, and Source Identification of Heavy Metals in River Sediments from Hai River and its Tributaries, Tianjin, China. *Water Air Soil Pollut.* 231 (2), 1–14. doi:10.1007/s11270-020-4404-6

Karbassi, A. R., Monavari, S. M., Nabi Bidhendi, G. R., Nouri, J., and Nematpour, K. (2007). Metal Pollution Assessment of Sediment and Water in the Shur River. *Environ. Monit. Assess.* 147 (1-3), 107–116. doi:10.1007/s10661-007-0102-8

Ke, X., Gui, S., Huang, H., Zhang, H., Wang, C., and Guo, W. (2017). Ecological Risk Assessment and Source Identification for Heavy Metals in Surface Sediment from the Liaohe River Protected Area, China. *Chemosphere* 175, 473–481. doi:10.1016/j.chemosphere.2017.02.029

Kumar, V., Sharma, A., Bhardwaj, R., and Thukral, A. K. (2018). Temporal Distribution, Source Apportionment, and Pollution Assessment of Metals in the Sediments of Beas River, India. *Hum. Ecol. Risk Assess. Int. J.* 24 (8), 2162–2181. doi:10.1080/10807039.2018.1440529

Lee, H.-S., Lim, S.-J., Lim, B.-R., Kim, H.-S., Lee, H.-S., Ahn, T.-U., et al. (2022). Spatiotemporal Evaluation of Water Quality and Hazardous Substances in Small Coastal Streams According to Watershed Characteristics. *Ijerph* 19 (2), 634. doi:10.3390/ijerph19020634

Li, X., Liu, L., Wang, Y., Luo, G., Chen, X., Yang, X., et al. (2014). Integrated Assessment of Heavy Metal Contamination in Sediments from a Coastal Industrial Basin, N. E. China. *Heavy Metal Contam. Water Soil*, 157–182. doi:10.1201/b16566-10

Liang, X., Li, J., Guo, G., Li, S., Gong, Q., Zhang, F. Y., et al. (2021). Evaluation for Water Resource System Efficiency and Influencing Factors in Western China: A Two-Stage Network DEA-Tobit Model. *J. Clean. Prod.* 328 (6), 129674–129828. doi:10.1016/j.jclepro.2021.129674

Liao, J., Chen, J., Ru, X., Chen, J., Wu, H., and Wei, C. (2017). Heavy Metals in River Surface Sediments Affected with Multiple Pollution Sources, South China: Distribution, Enrichment and Source Apportionment. *J. Geochem. Explor.* 176, 9–19. doi:10.1016/j.gexplo.2016.08.013

Liu, J., Li, S., Zhu, J., Li, Z. J., Wang, P., and Wu, N. D. (2018). Discussion on the Harm to Human Body by Several Kinds of Heavy Metal Elements and Preventive Measures. *China Resour. Compr. Util.* 9, 2528–2531. doi:10.3969/j.issn.1008-9500.2018.03.064

Lu, T., Lin, Q., Ke, C., Sun, R., Pan, D. B., Hong, X. Y., et al. (2013). Polycyclic Aromatic Hydrocarbons and Risk Assessment in the Surface Sediments from Lingdingyang, Pearl River Estuary. *J. Fish. Sci. China* 19 (6), 336–347. doi:10.3724/sp.j.1118.2012.00336

MacDonald, D. D., Ingersoll, C. G., and Berger, T. A. (2000). Development and Evaluation of Consensus-Based Sediment Quality Guidelines for Freshwater Ecosystems. *Archives Environ. Contam. Toxicol.* 39 (1), 20–31. doi:10.1007/s002440010075

Malik, N., Biswas, A. K., Qureshi, T. A., Borana, K., and Virha, R. (2009). Bioaccumulation of Heavy Metals in Fish Tissues of a Freshwater Lake of Bhopal. *Environ. Monit. Assess.* 160 (1-4), 267–276. doi:10.1007/s10661-008-0693-8

Mohiuddin, K., Alam, M., Ahmed, I., and Chowdhury, A. (2016). Heavy Metal Pollution Load in Sediment Samples of the Buriganga River in Bangladesh. *J. Bangladesh Agric. Univ.* 13 (2), 229–238. doi:10.3329/jbau.v13i2.28784

Muller, G. (1981). The Heavy Metal Pollution of the Sediments of Neckars and its Tributary: a Stocktaking. *Chem. Zeit* 105, 157–164. doi:10.3406/bch.1981.1937

Nasrabadi, T., Nabi Bidhendi, G., Karbassi, A., and Mehrdadi, N. (2010). Evaluating the Efficiency of Sediment Metal Pollution Indices in Interpreting the Pollution of Haraz River Sediments, Southern Caspian Sea Basin. *Environ. Monit. Assess.* 171 (1-4), 395–410. doi:10.1007/s10661-009-1286-x

Nemerow, N. L. (1991). *Stream, Lake, Estuary, and Ocean Pollution*. New York, NY: Van Nostrand Reinhold.

Niu, H., Deng, W., Wu, Q., and Chen, X. (2009). Potential Toxic Risk of Heavy Metals from Sediment of the Pearl River in South China. *J. Environ. Sci.* 21 (8), 1053–1058. doi:10.1016/s1001-0742(08)62381-5

Omwene, P. I., Öncel, M. S., Çelen, M., and Kobya, M. (2018). Heavy Metal Pollution and Spatial Distribution in Surface Sediments of Mustafakemalpaşa Stream Located in the World's Largest Borate Basin (Turkey). *Chemosphere* 208, 782–792. doi:10.1016/j.chemosphere.2018.06.031

Salati, S., and Moore, F. (2009). Assessment of Heavy Metal Concentration in the Khoshk River Water and Sediment, Shiraz, Southwest Iran. *Environ. Monit. Assess.* 164 (1-4), 677–689. doi:10.1007/s10661-009-0920-y

Shafie, N. A., Aris, A. Z., and Haris, H. (2014). Geoaccumulation and Distribution of Heavy Metals in the Urban River Sediment. *Int. J. Sediment Res.* 29 (3), 368–377. doi:10.1016/s1001-6279(14)60051-2

Shi, C. Y., Liang, M., and Feng, B. (2016). Average Background Values of 39 Chemical Elements in Stream Sediments of China. *Earth Sci.* 41 (2), 234–251. doi:10.3799/dqkx.2016.018

Siddiqui, E., and Pandey, J. (2019). Assessment of Heavy Metal Pollution in Water and Surface Sediment and Evaluation of Ecological Risks Associated with Sediment Contamination in the Ganga River: a Basin-Scale Study. *Environ. Sci. Pollut. Res.* 26, 10926–10940. doi:10.1007/s11356-019-04495-6

Sutherland, R. A. (2000). Bed Sediment-Associated Trace Metals in an Urban Stream, Oahu, Hawaii. *Environ. Geol.* 39 (6), 611–627. doi:10.1007/s002540050473

Tang, W. Z., Wang, L. S., and Shan, B. Q. (2015). Speciation Distribution of Heavy Metal in the Surface Sediments of a Typical Urban River (Liangshui River System). *Acta Scien Circum* 35 (12), 3898–3905. doi:10.13671/j.hjkxxb.2015.0021

Vaezi, A. R., Karbassi, A. R., and Fakhraee, M. (2015). Assessing the Trace Metal Pollution in the Sediments of Mahshahr Bay, Persian Gulf, via a Novel Pollution Index. *Environ. Monit. Assess.* 187 (10). doi:10.1007/s10661-015-4833-7

Wei, B., and Yang, L. (2010). A Review of Heavy Metal Contaminations in Urban Soils, Urban Road Dusts and Agricultural Soils from China. *Microchem. J.* 94 (2), 99–107. doi:10.1016/j.microc.2009.09.014

Wei, X., Han, L., Gao, B., Zhou, H., Lu, J., and Wan, X. (2016). Distribution, Bioavailability, and Potential Risk Assessment of the Metals in Tributary Sediments of Three Gorges Reservoir: The Impact of Water Impoundment. *Ecol. Indic.* 61, 667–675. doi:10.1016/j.ecolind.2015.10.018

Wu, S.-S., Han, R.-M., Yang, H., Bi, F.-Z., Wang, Q.-J., and Wang, Y.-H. (2017). Characterization of Metals in Surface Sediments from Xiaoyang River, Jiangsu, China. *Anal. Lett.* 50 (10), 1669–1690. doi:10.1080/00032719.2016.1241800

Xiao, R., Bai, J., Huang, L., Zhang, H., Cui, B., and Liu, X. (2013). Distribution and Pollution, Toxicity and Risk Assessment of Heavy Metals in Sediments from Urban and Rural Rivers of the Pearl River Delta in Southern China. *Ecotoxicology* 22 (10), 1564–1575. doi:10.1007/s10646-013-1142-1

Xu, R., Ni, S. J., Tuo, X. G., and Zhang, C. J. (2008). Progress in Nanoparticles Characterization: Sizing and Zeta Potential Measurement. *Particuology* 6 (2), 112–115. doi:10.3969/j.issn.1003-6504.2008.02.03010.1016/j.partic.2007.12.002

Yao, X., Luo, K., Niu, Y., Li, Y., and Ren, B. (2021). Ecological Risk from Toxic Metals in Sediments of the Yangtze, Yellow, Pearl, and Liaohe Rivers, China. *Bull. Environ. Contam. Toxicol.* 107 (1), 140–146. doi:10.1007/s00128-021-03229-0

Yongming, H., Peixuan, D., Junji, C., and Posmentier, E. (2006). Multivariate Analysis of Heavy Metal Contamination in Urban Dusts of Xi'an, Central China. *Sci. Total Environ.* 355 (1-3), 176–186. doi:10.1016/j.scitotenv.2005.02.026

Yuan, X., Zhang, L., Li, J., Wang, C., and Ji, J. (2014). Sediment Properties and Heavy Metal Pollution Assessment in the River, Estuary and Lake Environments of a Fluvial Plain, China. *CATENA* 119, 52–60. doi:10.1016/j.catena.2014.03.008

Zhang, W. J. (2020). Analysis of Design Strategy of Agricultural Tourism Park in Western Guangdong. *Contemp. Hortic.* (20), 67–68. doi:10.3969/j.issn.1006-4958.2020.20.033

Zhang, Y. Y. (2013). Reflections on the Economic Development of Guangdong and the West of Guangdong. *Sci-Tech Dev. Enterp.* 13, 7–9. doi:10.3969/j.issn.1674-0688.2013.13.003

Zhao, G., Ye, S., Yuan, H., Ding, X., and Wang, J. (2016). Surface Sediment Properties and Heavy Metal Pollution Assessment in the Pearl River Estuary, China. *Environ. Sci. Pollut. Res.* 24 (3), 2966–2979. doi:10.1007/s11356-016-8003-4

Zhong, X. Y., and Yang, H. W. (2001). Analysis of the Influence of Waste Water Entering Rivers in Guangdong Province on the Water Quality of Rivers in Our Province. *Guangdong Water Resour. Hydropower* (S1), 41–42+31. doi:10.3969/j.issn.1008-0112.2001.z1.017

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Environmental Science

TYPE Original Research
PUBLISHED 25 August 2022
DOI 10.3389/fenvs.2022.968693

OPEN ACCESS

EDITED BY
Rita Yi Man Li,
Hong Kong Shue Yan University, Hong Kong SAR, China

REVIEWED BY
Lifei Wang,
University of Toronto, Canada
Shujian Zhang,
Shenzhen University, China

*CORRESPONDENCE
Chuanfu Zang,
chuanfuzang@163.com

SPECIALTY SECTION
This article was submitted to Environmental Economics and Management, a section of the journal Frontiers in Environmental Science

RECEIVED 14 June 2022
ACCEPTED 25 July 2022
PUBLISHED 25 August 2022

CITATION
Li Y, Deng J, Zang C, Kong M and Zhao J (2022), Spatial and temporal evolution characteristics of water resources in the Hanjiang River Basin of China over 50 years under a changing environment. *Front. Environ. Sci.* 10:968693. doi: 10.3389/fenvs.2022.968693

Spatial and temporal evolution characteristics of water resources in the Hanjiang River Basin of China over 50 years under a changing environment

Yiting Li, Jinglin Deng, Chuanfu Zang*, Ming Kong and Jieni Zhao

School of Geography, South China Normal University, Guangzhou, China

In the formulation of scientific water resource management systems and strategies for improving water resource utilization efficiency, the effects of land use change and climate change on the temporal and spatial distribution of water resources cannot be overlooked. And the Hanjiang basin is the main source of water supply to eastern Guangdong and southwestern Fujian, China. However, the basin is experiencing a significant imbalance between water resource supply and demand. This study investigated the spatiotemporal distribution of water resources in the Hanjiang River Basin under changing environmental conditions using the Soil and Water Assessment Tool model. The findings reveal that, from 1980 to 2020, changes in water resources significantly varied among sub-basins, reflecting spatial heterogeneity. Moreover, sub-basins with severe land use changes showed significant changes in water resources. From 1970 to 2020, the water resources of each sub-basin changed with climate change, indicating temporal variability. Under the combined effects of land use change and climate change, the amount of water resources decreased and its spatial distribution changed dramatically. At the same time, the evolution of water resources under climate change was consistent with that under the combined influence of both land use change and climate change, indicating that climate change is the primary driver. The findings provide theoretical guidance for water resource research and management.

KEYWORDS

land use change, climate change, Hanjiang River Basin, SWAT model, water resource

1 Introduction

Water resources are one of the most important factors for maintaining ecosystem balance, human survival, and socio-economic development (Long et al., 2019). With the growth of the economy and society, human demand for water resources is growing, and the sustainable use of water resources has emerged as a major stumbling block to China's

growth (Piao et al., 2010). Changes in hydrological circumstances have potential long-term consequences on human civilization (Hall et al., 2014). Climate change and changes in land cover/use are frequently reported as two primary driving factors of hydrological change (Zhou et al., 2015) and are strongly associated with current and future water scarcity (Zhou et al., 2022). At long time scales, climate change plays a dominant role in influencing hydrological cycle processes in watersheds; at short time scales, land use change is one of the key factors influencing changes in hydrological elements in watersheds (Zhang et al., 2020a). Under the background of global warming, changes in rainfall, evaporation, runoff, and soil moisture will alter the global hydrological cycle, inducing a redistribution of water resources temporally and spatially (Xia et al., 2015). At the same time, continued population growth and highly intensive human activities also alter land cover conditions, which in turn induce changes in various hydrological processes such as infiltration and evaporation (Miller et al., 2014; Zang et al., 2015; Anand et al., 2018; Ross and Randhir, 2022). Many studies also suggest that land use change will negatively affect the sustainable use of water resources in the future (Shrestha et al., 2018; Ross and Randhir, 2022). The joint interaction of climate change and land use change has resulted in changes in terrestrial water cycle processes, redistribution of water resources on the ground, and even extreme hydrological events, further exacerbating the conflict between water supply and demand (Davis et al., 2015; Zang et al., 2015; Zhang et al., 2020b; Wu et al., 2021). In this context, the effects of climate change and human activities on water resources, as well as the characteristics of water resource evolution at the basin scale, should not be overlooked when developing scientific water resource management systems and strategies for improving water resource utilization efficiency.

As the impact of human activities and climate change on water resources is growing, the United Nations Educational, Scientific and Cultural Organization (UNESCO) Intergovernmental Hydrological Programme (IHP), International Geosphere Biosphere Programme (IGBP) and other large international scientific research programs are increasingly focusing on the impact of these two factors to establish a scientific foundation for the transformation and rational use of water resources (Sivakumar, 2011). Many researchers in China and worldwide have studied the effects of changing scenarios (such as land use change and climate change) on water resources under this setting. Piao et al. (2010) investigated the effects of climate change on water resources and agriculture in China by analyzing climate change trends and water resource trends (2010). Lotz et al. (2018) quantified the hydrological response of the Dongting Lake watershed in China to land use change using the Soil and Water Assessment Tool (SWAT) model, and they found that the transition from agricultural land to forest land reduced surface runoff and total water, while increasing evapotranspiration, subsurface flow, and groundwater flow (2018). Kundu et al. (2017) investigated the effects of climate and land use change on future water balance, and concluded that climate change has a stronger influence on water production, whereas land use change has a stronger influence on evapotranspiration (2017). In general, most studies have focused on changes in water resources within a single environment. In terms of research methods, many scholars have studied the use of various hydrological models to assess the combined effects of land use change and climate change. Among these, the SWAT model is an integrated tool for multidisciplinary studies at the regional scale in diverse physiographic and climatic conditions. It is also one of the most widely used watershed-scale models in the world today (Krysanova and White, 2015; Souffront Alcantara et al., 2019). Therefore, more in-depth quantitative analyses of the spatiotemporal evolution characteristics and driving mechanisms of water resources at the basin scale are required with the help of SWAT models, considering both human activities and climate change. Such analyses are critical for water resource planning and management.

Apart from the Pearl River Basin, eastern Guangdong and southern Fujian in China rely significantly on the Hanjiang River for water. With a population of over 10 million people and a combined regional gross domestic product (GDP) of over 190 billion Yuan, the Hanjiang River Basin is the second-largest basin in Guangdong (Li et al., 2021), and occupies a very important position in the economic and social development of Guangdong. In the context of global climate change, changes in the runoff of the Hanjiang River Basin are becoming increasingly significant, with frequent floods in local areas in the flood season and more severe water shortages or even severe droughts in the non-flood season (Feng and Hu, 2021). For example, the Hanjiang River Basin suffered the most severe drought in 60 years in 2017. As precipitation is the primary source of water in the Hanjiang River Basin, climate change directly affects the spatiotemporal distribution of water resources in the basin (Li et al., 2021). With the implementation of the reform and opening up of China since 1978, the Hanjiang River Basin has experienced rapid economic development, and with urbanization, land use types in the Hanjiang River Basin have shifted substantially (Zhang, 2007), which affect hydrological processes such as runoff confluence, evaporation, and dispersion. Major reservoirs in the Hanjiang River Basin have limited water storage capacity, with their total water storage capacity accounting for only about 5% of the water resources in the entire basin, which is far below the national average 12% (Li et al., 2021). Therefore, the basin has limited resistance to extreme hydrological events.

This study investigated the spatiotemporal distribution of water resources in the Hanjiang River Basin considering changes in human activities and climatic conditions. The effects of climate change and land use type change on both the temporal and spatial variability of water resources in the Hanjiang River Basin

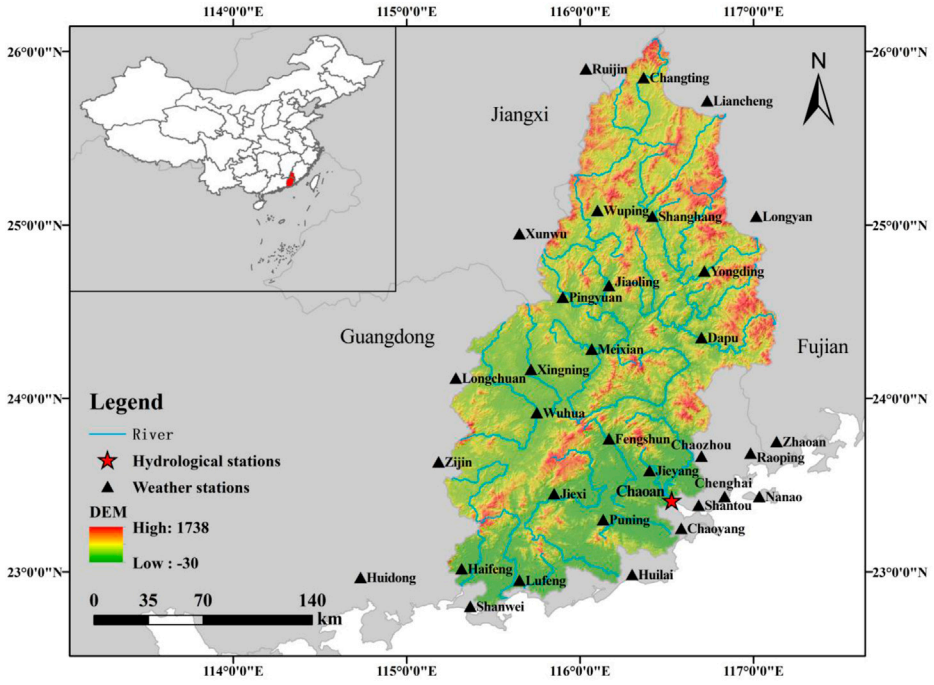


FIGURE 1
Location and distribution of meteorological stations in the Hanjiang River Basin, China.

under several scenarios were quantitatively analyzed using the SWAT model. Furthermore, the geographical distribution pattern of water resources was investigated at the basin scale. The geographical distribution and usage of water resources were combined to obtain a theoretical foundation and scientific basis for the integrated management, water security, ecological security, and long-term economic growth of the Hanjiang River Basin. The findings will promote rational scheduling and management of water resources in the basin, modernization of the water governance system and governance capacity in the basin, and high-quality, long-term economic growth in the basin.

2 Study area

The Hanjiang River Basin is mainly located in eastern Guangdong and southwestern Fujian, China. It covers some counties and cities in Guangdong, Fujian, and Jiangxi provinces (Figure 1). Following the Pearl River Basin, this basin is the second-largest river basin in Guangdong Province. The area of the basin is 30,112 km^2, of which 13,929 km^2 (46.3%) is covered by the Meijiang River, 11,802 km^2 (39.2%) by the Ting River, 3,346 km^2 (11.1%) by the Hanjiang River, and 1,035 km^2 (3.4%) by the Hanjiang River Delta (below Chaoan Station). The Hanjiang River Basin is located in the subtropical southeast Asian monsoon region, which has subtropical climate. The climate is hot and humid, with abundant rainfall and frequent rainstorms. The annual average rainfall is approximately 1,600 mm. Precipitation shows uneven annual distribution and wide spatial variability, controlled by the topography.

The variability of annual rainfall in the Hanjiang River Basin causes large disparities in runoff in the dry and flood seasons, with the average discharge in the flood season being 2.7 times that of the dry season. The total annual runoff is 24.5 billion m^3, the average depth of yearly runoff is 600–1,200 mm, the average annual runoff coefficient is 0.51, and the yearly runoff distribution is unbalanced. Approximately 80% of the yearly runoff occurs from April to September. The spatial distribution of runoff is roughly consistent with the rainfall distribution. The topography of the Hanjiang River basin is predominantly mountainous and hilly. The mountainous area accounts for 70% of the total basin area, mostly distributed in the northern and central regions, and the general elevation is above 500 m. Hilly areas account for 25% of the total basin area, mostly distributed in the Meijiang River Basin and other major tributaries, with a general elevation below 200 m. Plain areas account for 5% of the total basin area, mostly in the delta downstream of the Hanjiang River, with a general elevation below 20 m. The main land use types are forest, cultivated land, grassland, and construction land, covering about 97% of the basin area.

TABLE 1 Data used to construct the Soil and Water Assessment Tool for the Hanjiang River Basin, China.

Data type	Data source	Data description
Digital elevation Model (DEM) data	Chinese Academy of Sciences data cloud (http://www.csdb.cn/)	90 m resolution
Land use data	Land use data obtained through remote sensing interpretation, and Remote sensing data obtained from the official website of the US Geological Survey (http://earthexplorer.usgs.gov)	Five periods in 1980, 1990, 2000, 2010, 2020; 30 m resolution
Soil data	Harmony World Soil Database, (https://www.fao.org/soils-portal/soil-survey/soil-maps-and-databases/harmonized-world-soil-database-v12/en/);	1:1 million
Meteorological data	National meteorological science data sharing service platform (http://data.cma.cn/site/index.html) and some district and county meteorological bureaus of Guangdong Province, Fujian Province, and Jiangxi Province.	Daily-scale data from 1969 to 2020
Runoff data	Hanjiang River Basin Administration	Chaoan Hydrological Station. 1981–1995; 1996–2010

3 Materials and methods

3.1 Materials

It is necessary to collect and preprocess relevant data of the Hanjiang River Basin and finally import the model database. The construction of the SWAT model required the import of sub-catchment data along with other data. The required data include Digital Elevation Model Data (DEM), land use data, soil data, and meteorological data (Table 1).

It is an important step to make sure the variables meet the underlying assumptions of the algorithms before conducting any statistical analyses. Therefore, following data collection, many types of data must be processed. Considering the Digital Elevation Model Data and Land-Use and Land-Cover Change Data are consistent with the target data properties required for modeling, we do less transformation processing on these data. Contrarily, conversion processing is necessary for the original soil and meteorological data since their characteristics differ from those needed to create the model. First, the raw Digital Elevation Model Data was processed with Arcgis software. Second, the land use categories map was reclassified in accordance with the watershed's real circumstances using the Land-Use and Land-Cover Change classification method, which corresponds to CropLu and UrbanLu, and ultimately the five reclassified land use maps were obtained. Third, the processing and integration of soil data is required for the building of the hydrological model. As a result, we gathered soil type and property data from the World Soil Database and then utilized the Soil-Plant-Atmosphere-Water (SPAW) program created by Washington State University to compute soil bulk density, effective field capacity, saturated hydraulic conductivity, and other parameters. Forth, for this study, observational data from 32 meteorological stations in the Han River watershed from 1969 to 2020 must be processed. In order to create the weather file needed for modeling, the data must first be sorted and validated before being fed into the weather generator.

3.2 Hydrological model: Soil and water assessment tool

Soil and Water Assessment Tool (SWAT) utilizes meteorological, hydrological, and management data to calculate the internal circulation of each hydrologic response unit (HRU) individually. The model then groups them into sub-basins and connects them organically through the river network to approximate surface catchments. Therefore, the model can not only evaluate the distribution of water resources but also identify and simulate non-point source pollution in critical areas (Arnold et al., 1998). Currently, the SWAT model has been widely used to simulate and assess the impact of land use change on watershed hydrology (Gassman et al., 2007). Therefore, the hydrological module of this model was selected to simulate hydrological processes in the Hanjiang River Basin.

SWAT simulates two types of hydrological processes: the hydrological cycle and the confluence phase. The water balance equation is the foundation of the hydrological cycle (Osei et al., 2019):

$$SW_t = SW_0 + \sum_{i=1}^{t}\left(P_{day} - Q_{surf} - E_a - W_{seep} - Q_{gw}\right) \quad (1)$$

In Eq. 1, SW_t is the final soil water content (mm); SW_0 is the pre-soil water content (mm); t is the time step (d); P_{day} is precipitation on day i (mm); Q_{surf} is surface runoff on day i (mm); E_a is evapotranspiration on day i (mm); W_{seep} is soil infiltration and lateral flow on day i (mm); and Q_{gw} is subsurface runoff on day i (mm).

In this study, ArcSWAT version 2012 was used for the construction and running of the SWAT model after data pre-processing. The watershed was divided into 48 sub-basins with topographic data and minimum catchment area (CSA) thresholds from the Digital Elevation Model Data of the Hanjiang River Basin. Secondly, the sub-basins were further divided into 931 HRUs within each sub-basin based on a combination of specific land use and soil type thresholds to

ensure that the entire sub-basin area was modeled. Finally, 32 meteorological stations and related files were written for the period 1969–2020. After selecting the suitable calculation technique and output items, the commands to run the model were executed.

Before calibrating the model, the parameters that have the most influence on the streamflow simulation need to be estimated by sensitivity analysis. Since ArcSWAT contains a large number of hydrological parameters and all of them may have an impact on the model effects, important input parameters for the simulation need to be identified (Uniyal et al., 2015). After the SWAT model run is completed, the model needs to be calibrated and validated to evaluate its applicability. SWAT-CUP (Calibration and Uncertainty Program) is a stand-alone program developed to calibrate SWAT (Abbaspour et al., 2007). The procedure evaluates the sensitivity of model parameters using multiple regression values of the objective function based on Latin hypercube sampling, integrating the SUFI2 (Sequential Uncertainty Fitting) algorithm, PSO (Particle Swarm) algorithm, GLUE (Maximum Likelihood) algorithm, ParaSol algorithm, and MCMC (Monte Carlo) algorithm. The SUFI2 algorithm is a method for calculating the objective function by randomly generating a set of parameters substituted into SWAT by the Latin-Hypercube simulations (Abbaspour et al., 2015).

Especially in hydrology, correlation and correlation-based measures such as the coefficient of determination and the Nash-Sutcliffe coefficient have been widely used to evaluate the "goodness-of-fit" of hydrologic and hydroclimatic models (Legates and McCabe, 1999). Coefficient of determination is a useful metric for a general interpretation of the efficiency of a model or a part (Nash and Sutcliffe, 1970), as it is widely used not only in hydrology but also for modeling and forecasting purposes in domains related to ecology, agriculture, and climatology (Onyutha, 2022). Its value ranges from 0 to 1. The higher value of it is associated with an effective model. However, it's value can be low and high for an accurate and imperfect model, respectively. Meanwhile, it does not quantify a model's bias (Onyutha, 2022). Therefore, the coefficient of efficiency such as the Nash-Sutcliffe coefficient improve the coefficient of determination for model evaluation purposes, as it is sensitive to differences in the observed and model- simulated means variances (Legates and McCabe, 1999). For that reason, modelers, especially hydrologists, tend to agree that the NSE is the optimal version of R-squared in general (Nash and Sutcliffe, 1970; Kvålseth, 1985). The value range of Nash-Sutcliffe coefficient is $(-\infty, 1]$, when $E_{ns} = 1$, the simulation effect is the best; when $E_{ns} < 0$, it indicates that the simulated and actual values are poorly fitted (McCuen et al., 2006). Many scholars believe that the simulation of the model is considered good when $E_{ns} > 0.75$, good when E_{ns} is between 0.36 and 0.75, and poor when $E_{ns} < 0.36$ (Santhi et al., 2001; McCuen et al., 2006). R^2 and E_{ns} are calculated as follows.

$$R^2 = \frac{\left[\sum_{i=1}^{n}\left(Q_{Oi} - \overline{Q_O}\right)\left(Q_{mi} - \overline{Q_m}\right)\right]^2}{\sum_{i=1}^{n}\left(Q_{Oi} - \overline{Q_O}\right)^2 \sum_{i=1}^{n}\left(Q_{mi} - \overline{Q_m}\right)^2} \quad (2)$$

$$E_{ns} = 1 - \frac{\sum_{i=1}^{n}(Q_{mi} - Q_{Oi})^2}{\sum_{i=1}^{n}\left(Q_{Oi} - \overline{Q_O}\right)} \quad (3)$$

In the equation, Q_{mi} is the simulated runoff series; Q_{Oi} is the measured runoff series; $\overline{Q_O}$ is the arithmetic mean of the measured runoff series; $\overline{Q_m}$ is the arithmetic mean of the simulated runoff series; n is the number of simulated periods.

In this study, the Chaoan station, an important hydrological station in the Han River basin, was selected to calibrate the model for the monthly runoff in the Han River basin from 1969 to 2020 with a simulation step of months. Among them, 1969–1980 is set as the pre-calibration period, 1981–1995 as the calibration period, and 1996–2010 as the validation period. The SWAT model was run with the important complementary tool SWAT-CUP which provides the SUFI2 algorithm to perform parameter sensitivity analysis and rate and validate the sensitive parameters to make the model accurate.

3.3 Water resource statistics

3.3.1 Calculation of water resources

When calculating water resources, this study did not simply count the surface runoff of each sub-basin. Instead, the sum of seepage, surface runoff, and lateral flow in the root zone was calculated using SWAT model simulations based on the main components and functions of water resources and combined with the Hanjiang River basin's characteristics.

$$W = \sum_{i=1}^{n}(\mathrm{PERC} + \mathrm{SURQ} + \mathrm{LATQ}) * S_i * 1000 \quad (4)$$

In Eq. 4, i represents the i-th HRU and n represents the total number of HRUs; W is the total water resources (m^3); PREC is seepage in the root zone during the time step (mm); SURQ is the surface runoff generated by the hydrologic response unit (HRU) within a certain period of time (mm); LATQ is lateral flow generated in the HRU during timestep (mm).

3.3.2 The relative change rate

Under various scenarios, the relative rate of change of water resources in the basin might reflect the change of water resources (Zang et al., 2012). The relative change rate (RCR) of water resources in different periods can be expressed as:

$$RCR = \frac{W_2 - W_1}{W_1} \times 100\% \quad (5)$$

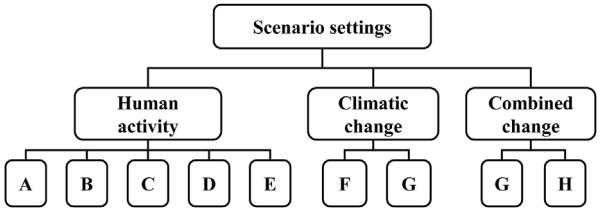


FIGURE 2
Scenarios and conditions considered for analysis.

TABLE 2 Periods of land use and meteorological data considered in the scenarios.

Scenario	Year	
	Land use data	Meteorological data
A	1980	1970–2020
B	1990	1970–2020
C	2000	1970–2020
D	2010	1970–2020
E	2020	1970–2020
F	2020	1971–1980
G	2020	2011–2020
H	1980	1971–1980

In Eq. 5, RCR is the relative rate of change in water resources, W is the water resource for each scenario, and W1 and W2 are the water resources in the previous and latter period, respectively.

3.4 Scenario settings

After completing model evaluation, eight scenarios were set up (Figure 2; Table 2) to quantitatively analyze the changes in the spatial and temporal distribution of water resources in the basin under human activities and climate change. Based on these scenarios, the effects of land use variability, climate change on water resources were assessed, as well as their combined impacts.

The periods of data used in eight simulation experiments are as follows. In Scenarios A, B, C, D, and E, meteorological data of 1970–2020 were used. Regarding land use data for these scenarios, data of 1980 were used for Scenario A, 1990 for Scenario B, 2000 for Scenario C, 2010 for Scenario D, and 2020 for Scenario E. For Scenario F, land use data of 2020 and meteorological data of 1971–1980 were used. For Scenario G, land use data of 2020 and meteorological data of 2011–2020 were used. For scenario H, land use data of 1980 and meteorological data of 1971–1980 were used. Scenarios A–E were compared to analyze the impact of land use change on the spatial and temporal distribution of water resources. By comparing scenarios F–G, the impact of climate change on the spatial and temporal distribution of water resources was investigated. Scenario G was compared with Scenario H to assess the impact of the two combined scenarios on the spatial and temporal distribution of water resource.

4 Results

4.1 Validation results

The calibration and validation of the SWAT model simulations of the downstream hydrographic station (Chaoan Station) provided satisfactory results (Figure 3). The correlation coefficient R^2 and the Nash-Sutcliffe efficiency coefficient were used to evaluate the SWAT model simulation results. For the calibration period, R^2 = 0.95 and E_{ns} = 0.94, and for the validation period, R^2 = 0.95 and E_{ns} = 0.93. The accuracy of the runoff simulation values for the calibration and validation periods was high, and the high R^2 indicates that the simulated and observed runoffs are highly correlated. Overall, the model simulation results were satisfactory, indicating the applicability of the model to the analysis of the regional and temporal distribution of water resources in the Hanjiang River Basin under changing conditions.

4.2 Impact of land use change

The flow of water in the Hanjiang River Basin, as well as the spatial and temporal distribution of water resources, has changed as a result of human activity. Land use categories in the basin have shifted as a result of human activity, and the nature of the substratum affects water cycle processes, resulting in changes in water resources. Under changing land use types from 1980 to 2020, the magnitude of changes in water resources and the magnitude of changes in the sub-basins of the Hanjiang River Basin fluctuated.

The most dramatic change in water resources was observed between 1980 and 1990 (Scenario A to Scenario B), as shown in Figure 4. Water resources in all sub-basins increased by an average of 36.20%. The increase in construction land and decline in farmland and forest area are the primary causes of the significant increase in water resources in the midstream and downstream areas. The loss of cropland and woodland eventually resulted in an increase in surface runoff and an increase in the total water resources, whereas the increase in the area of impervious surface negatively influences water infiltration.

No significant change in water resources was observed between 1990 and 2000 (Scenario B to Scenario C). A decline in water

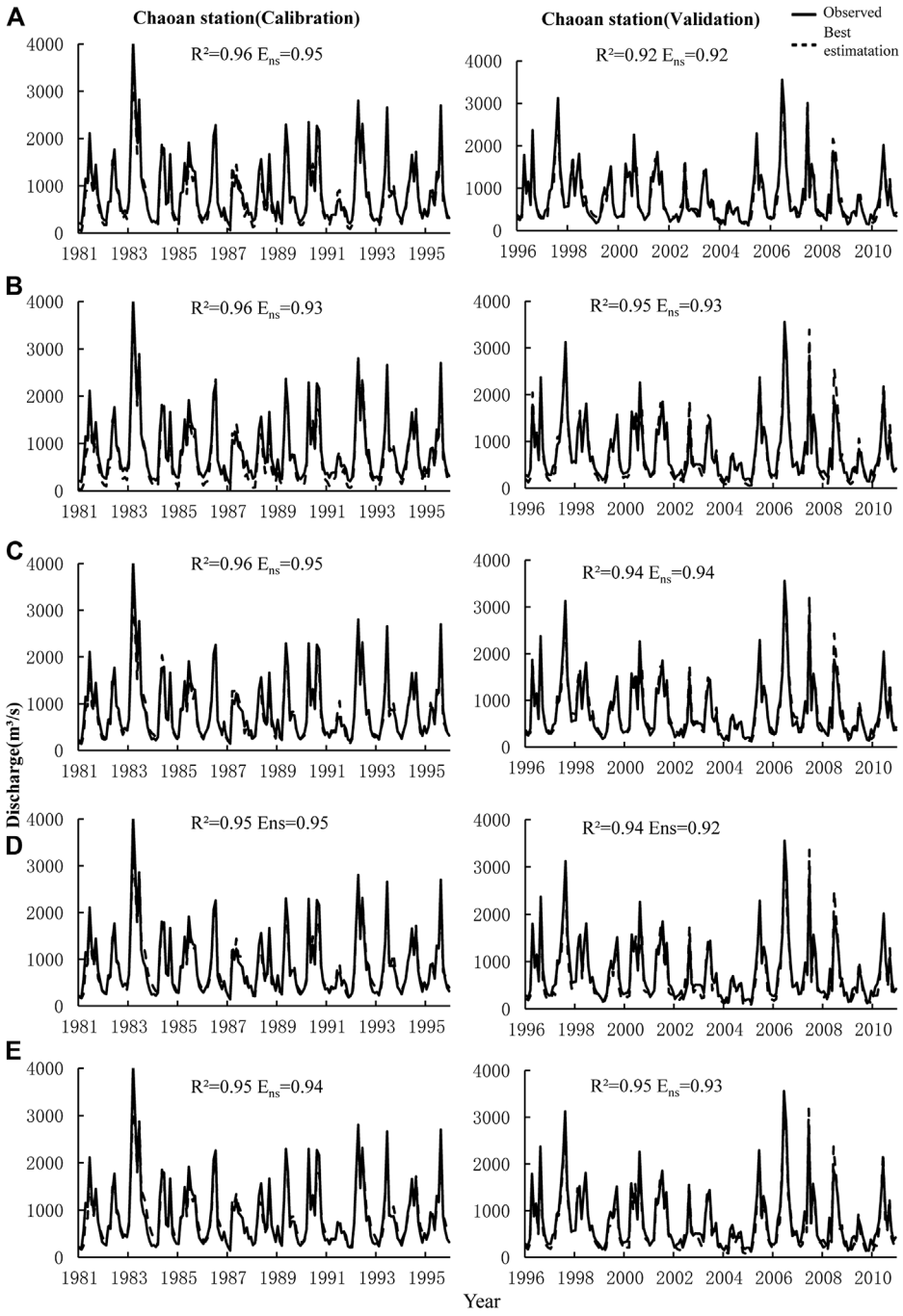


FIGURE 3
Under **(A–E)** scenarios, comparison of the soil and water assessment tool (SWAT) models simulations results to observed data for the Chaoan station over the calibration (1981–1995) and validation periods (1996–2010) in the Hanjiang River Basin, China (expressed as 95% prediction uncertainty band).

resources was observed in all sub-basins, but the decrease was small, with an average change of 15.01%. Although the amount of land used for construction has increased in some sub-basins, the area of forest land has also increased. Moreover, the increase in forested land is more notable than the increase in land used for construction, reducing the flow of water.

The change in water resources in each sub-basin from 2000 to 2010 (Scenarios C to D) was essentially the same as

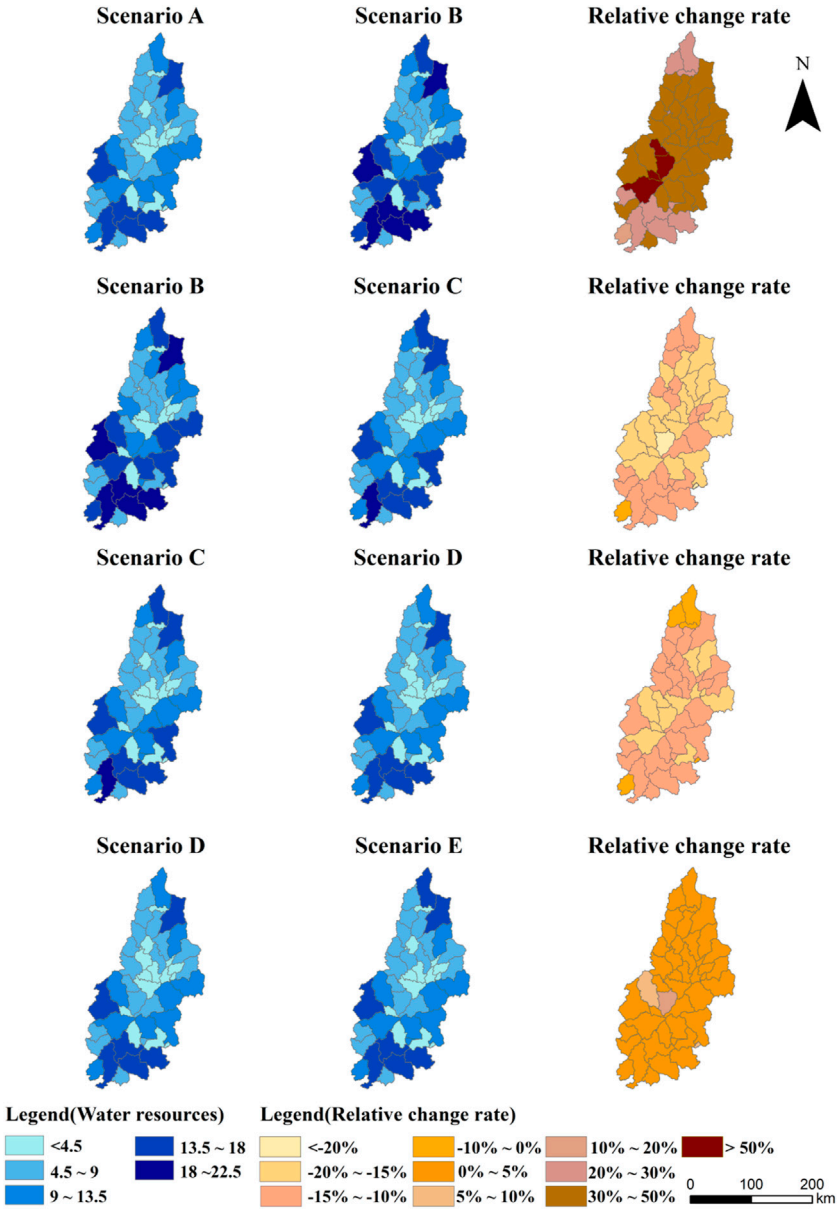


FIGURE 4
Water resources in the Hanjiang River Basin in different human activity scenarios. The impacts were assessed with relative change rate.

that from 1990 to 2000 (Scenarios B to C), although with a comparatively large reduction. The area of forest land increased in the downstream western sub-basin and the area of mixed wetlands increased in the upstream eastern sub-basin, with relatively considerable water resources losses in these areas.

Water resources increased in all sub-basins from 2010 to 2020 (Scenarios D to E), but not significantly. It increased considerably in individual sub-basins owing to the significant expansion in built-up land area in this region. In these scenarios, urbanization altered the spatial distribution of water while increasing water resources in all sub-basins.

$$*Relative\ change\ rate = \frac{W2 - W1}{W1} \times 100\%$$

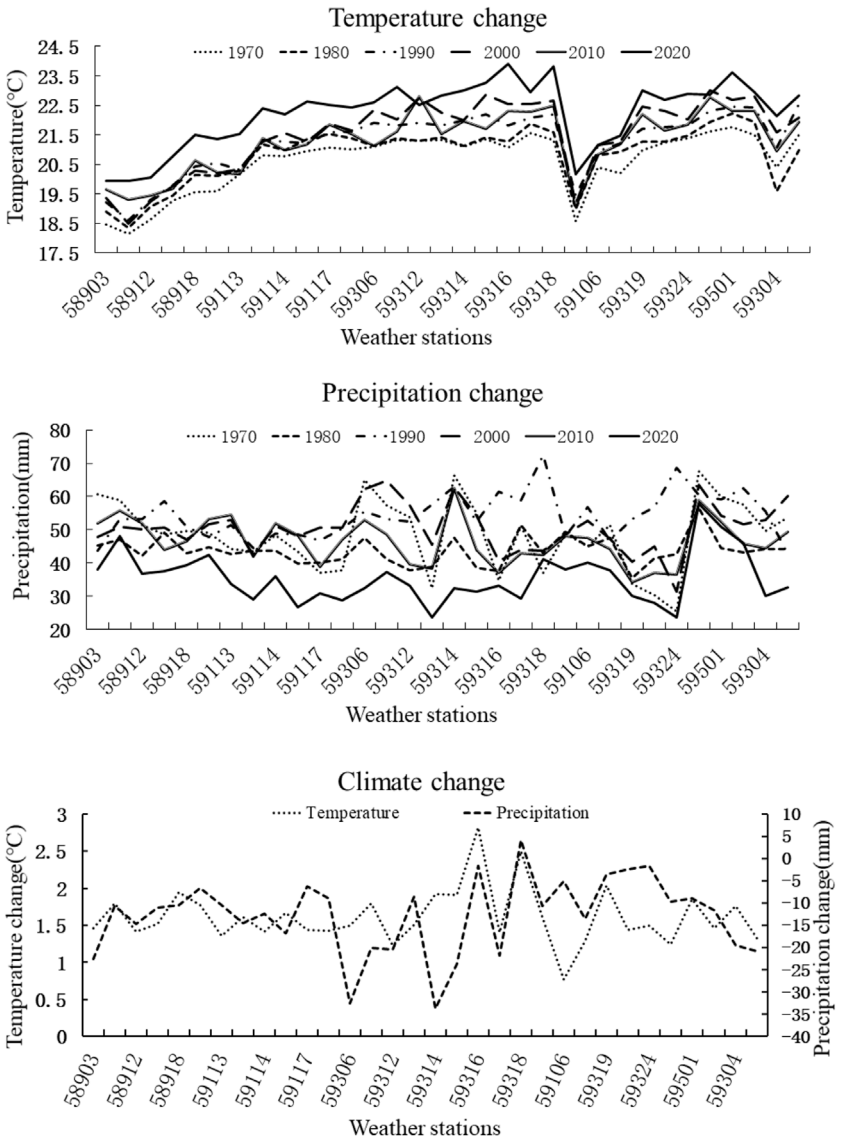


FIGURE 5
Climate change at meteorological stations in the Hanjiang River basin from 1970 to 2020.

In this equation, W1 and W2 are the water resources in the previous and latter period respectively.

4.3 Impact of climate change

According to variations in annual average precipitation and annual average temperature throughout the basin and in each sub-basin from 1970 to 2020, temperature continuously increased every year, while precipitation exhibited significant temporal and spatial variability (Figure 5). After analyzing changes in water resources in the basin under climate change scenarios, changes in water resources were found to be concentrated in the midstream and downstream areas, as well as the lower reaches (Figure 6). Water resources increased significantly in the midstream and downstream areas of the sub-basin, with an average rate of change of approximately 20%. In contrast water resources decreased in the eastern half of the downstream sub-basin, with an average rate of change of approximately 18%. These variabilities are mostly attributable to changes in precipitation and temperature. Temperature exhibited a continuous increase on a basin-wide scale, while precipitation in the middle and upper reaches increased and subsequently decreased, as a result of which water resources

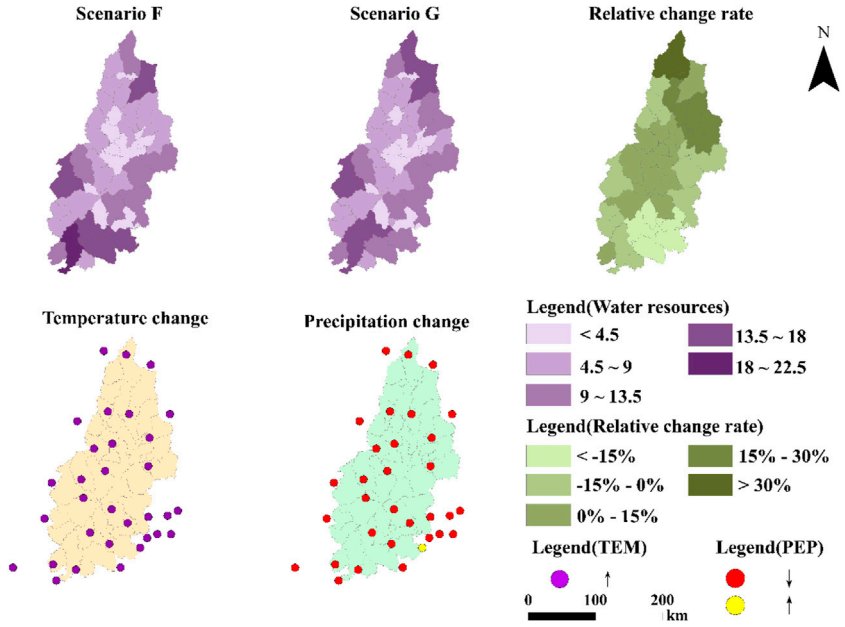


FIGURE 6
Climate change in the Hanjiang River Basin and changes in water resources in the sub-basins under the influence of climate change. The impacts were assessed with relative change rate.

increased in some of the midstream sub-basins. From 1970 to 2020, the eastern sub-basin of the lower reaches experienced a decline in water resources due to a decrease in precipitation and an increase in temperature. On this basis, both temperature and precipitation can be ascertained to influence the spatial and temporal distribution of water resources in the Hanjiang River Basin. Specifically, temperature exhibited a rising trend from 1970 to 2020, whereas precipitation exhibited variable increase with uneven spatial distribution. Moreover, precipitation widely varied among sub-basins, which may lead to increased differences in the spatial and temporal distribution of water resources.

$$*Relative\ change\ rate = \frac{W2 - W1}{W1} \times 100\%$$

In this equation, W1 and W2 are the water resources in the previous and latter period respectively.

4.4 Impact of land use change and climate change

The total water resources in the basin decreased by 8.92 million m^3 as a result of human activities and climate change, but total evapotranspiration increased by 9.09 million m^3. According to the water balance formula, the overall change in total water resources and total evapotranspiration in the Hanjiang River Basin were almost equivalent. According to water change rate statistics for each sub-basin between 1980 and 2020, water resources exhibited a rising trend in 31 sub-basins, accounting for 65% of the entire basin, and it exhibited a declining trend in 16 sub-basins, accounting for 35%. The rates of change of water resources in each sub-basin from 1980 to 2020 are compared in Figure 7. The figure reveals that the spatial distribution of water resources in the Hanjiang River Basin has dramatically changed over the years. Water resources dramatically changed in the middle and upper reaches of the basin, leading to considerable increases in overall water resources, with an average rate of change of more than 16%. Water resources dramatically increased in most of the central sub-basins, with some sub-basins experiencing a rate of change of more than 30%. Human activities changed the land use type between 1980 and 2020, as reflected by the following changes. With the increase in the area of building land, surface water infiltration was decreased and the water resources increased. The decrease in water resources in the southern sub-basin was more noticeable, with an average rate of change of more than 15%. According to the above finding, this decrease in water resources is primarily attributable to climate change, as the temperature of the region continuously increased from 1980 to 2020, whereas precipitation decreased.

Analyzing the changes in the basin's total water resources under the combined change scenario reveals a decline in the basin's total water resources, which implies that the degree of the

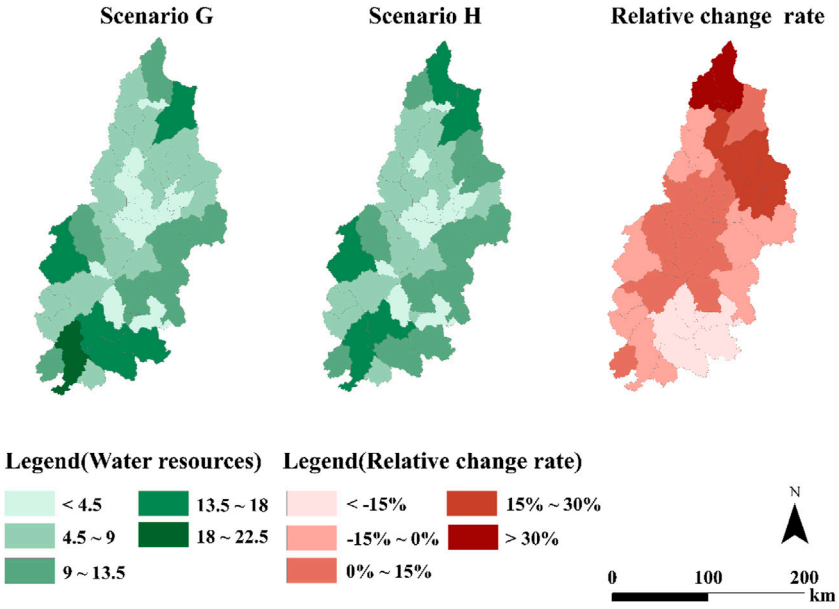


FIGURE 7
Changes in water resource at the sub-basin scale under the influence of combined conditions.

influence of climatic conditions on water resource changes may be greater than that caused by changes in land use. Additionally, each sub-basin's water resources are roughly comparable to those in each sub-basin in a scenario of climate change, suggesting that the climate may have been the primary driver of water resource evolution in the Hanjiang River basin.

$$*Relative\ change\ rate = \frac{W2 - W1}{W1} \times 100\%$$

In this equation, W1 and W2 are the water resources in the previous and latter period respectively.

5 Discussion

Taking the Hanjiang River Basin as the research object, this study investigated the spatial and temporal distribution characteristics of water resources under changing conditions, and quantitatively analyzed the influence of different conditions on water resources. This study fills a gap in existing studies and provides a theoretical basis for the planning and management of water resources in the Hanjiang River Basin. Two main findings can be deduced from the results: first, anthropogenic land use change has significantly altered the geographical and temporal distribution patterns of water resources; second, the temporal variability of water resources is more prominent with climate change, and the evolution characteristics of water resources under climate change are more consistent with those under the combined influence of both land use change and climate change.

Changes in land use type driven by human activities were found to directly alter the geographical and temporal distribution of water resources in the Hanjiang River Basin. Between 1980 and 2020, the regional and temporal distribution of water resources in the Hanjiang River Basin changed dramatically, but the overall change in the amount of water resources was very small. Changes in land use can affect water production by altering hydrological processes such as evapotranspiration and soil moisture dynamics (Sterling et al., 2013; Anand et al., 2018; Zhang et al., 2018). The land cover pattern within the Hanjiang River Basin has changed dramatically with the development of the social economy since the reform and opening up of China starting from 1978 (Zhang, 2007). The area of construction land within the basin increased significantly from 1980 to 2020, with an average change rate of more than 200%. The impervious ground of construction land to some extent reduces the production sink time and increases the flood peak, leading to an increase in regional water production (Miller et al., 2014; Strohbach et al., 2019). It is noteworthy that forest land is the primary land use type in the Hanjiang River Basin (Zhang, 2007). Changes in the extent of forest land are also important to the development of water resources in the basin, and the loss of woodland area negatively influences the processes of vegetation retention, lateral flow, and evapotranspiration, resulting in an increase in runoff (Brown et al., 2005; Foley et al., 2005; Sajikumar and Remya, 2015). Consequently, the spatial distribution pattern of water resources within the Hanjiang River Basin was altered. Therefore, the major reasons for the change in the spatial and temporal distribution of water resources are the growth in building land and the reduction in forest land caused by human activities.

Regarding the evolution of water resources under climate change, changes in climatic conditions from 1970 to 2020 induced significant temporal variability of water resources in the Hanjiang River Basin. In addition, the characteristics of changes in water resources were similar to those under combined conditions. Climate change (i.e., rising temperatures and changing rainfall patterns) is anticipated to have a major influence on the hydrological cycle by altering the spatial and temporal distribution of water cycle elements such as precipitation, evaporation, runoff, and soil moisture, which are predicted to affect the reallocation of water resources (Wu et al., 2012; Wang et al., 2013; Zang and Liu, 2013). Precipitation in the Hanjiang River Basin fluctuated downward from 1970 to 2020 while temperature continuously increased, resulting in increased evapotranspiration (Li and Fang, 2021). Consequently, the average annual surface runoff, baseflow, groundwater recharge, and thus total water quantity in the basin decreased. Furthermore, the characteristics of the changes in water volume varied with climate change in each sub-basin from 1970 to 2020, and the changes in water volume were more significant in sub-basins with more fluctuating climatic conditions, indicating that climate change is the main driver of natural temporal changes in the total water volume and water distribution in the basin. Furthermore, changes in the water volume in each sub-basin under climate change were similar to those under combined conditions, indicating that climate change is currently the most influential factor controlling the spatial and temporal evolution of water resources in the Hanjiang River Basin.

Compared to previous research findings, the current study presents similar but distinct findings. From a common point of view, climate change affects the evolution of water resources over time, and the effects of climate change are stronger than that of land use (Hagemann et al., 2013; Anand et al., 2018; Shrestha et al., 2018; Ross and Randhir, 2022). However, because anthropogenic land use change does not always imply a change in total water resources but rather changes in the distribution pattern of water resources, its impact is more visible at the basin scale (Zang et al., 2015; Anand et al., 2018; Ross and Randhir, 2022). In terms of various factors, some scholars have explored future climate change trends and their impact on water resources using climate change models and concluded that water resources in southern China will decrease significantly (Hagemann et al., 2013; Ljungqvist et al., 2016). In contrast, the results of this study suggest that the decreasing trend of future water resources in the Hanjiang River Basin is not significant. The difference between the findings of this study and previous studies can be attributed to the difference in the approaches employed. In this study, the evolution of water resources under climate change conditions was investigated with the scenario setting approach, and because the number of scenarios established is limited, the difference in the time scales of the climate change scenarios might lead to varying results.

We provide the following recommendations in light of the findings of this research and the situation of the water resources in the Hanjiang River Basin. First, strengthen restrictions on the use of water resources, such as strengthening oversight of water resources of crucial sub basins, and severely limit activities that have a negative impact on water resources. Second, we should standardize water consumption and promote water conservation at the source in important sectors of the economy, including agriculture and industry. Third, improve the functional control and allocation of water resources, as well as the management of water intake licenses and the management of planned and quota water consumption; Stop illegal groundwater exploitation and strengthen groundwater management and protection. The above ideas enable the sustainable expansion of the economy, society, and ecological environment by following the law of change to the management of water resources in a scientific and efficient manner.

Although the spatiotemporal evolution of water resources was comprehensively analyzed, some limitations still remain and they need to be further investigated. First, there are uncertainties in the simulation of the hydrological model (SWAT). In this study, monthly runoff data from the Chaoan station covering 1981–2010 were used to calibrate and validate the model, and the spatial location of the hydrological station will affect the model simulation performance. Second, the climate change scenarios and integrated scenarios in this study are relatively simple. Therefore, more climate change scenarios are required to analyze the specific impact mechanisms of climate change. Third, only a few indicators were considered to examine land use change induced by human activities. Moreover, the influence mechanism of each land use type on the evolution of water resource is still unknown. To clarify the influence mechanism of each land use type on water resources, subsequent studies should consider multiple indicators to evaluate land use change. Fourth, other factors such as topography, geochemistry, and the dynamic evolution of species communities may also have an impact on how watershed water resources evolve, making it possible to conduct a more comprehensive study of this evolution in the future based on the aforementioned aspects.

6 Conclusion

In this study, the SWAT model was calibrated and validated using monthly runoff data covering 1969–2020 from the Chaoan hydrological station in the Hanjiang River Basin, and eight types of scenarios were set up to quantitatively analyze the influence of anthropogenic land use change and climate change on the regional and temporal evolution of water supply in the Hanjiang River. From the results, the following conclusions can be drawn.

Under the influence of human activities, the distribution of water resources in the Hanjiang River Basin exhibited spatial heterogeneity between 1980 and 2020, which is mostly expressed in the varied evolution of water resource among sub-basins. Rapid development in the basin has had an impact on the water cycle, aggravating disparities in water distribution across the basin. Changes in land use type due to human activities bring about changes in hydrological processes within the basin, thus

changing the spatial and temporal distribution of water resources. These results suggest that land use changes are the primary factor controlling the spatial evolution of water resources.

Under the climate pattern of continuous rise in temperature and fluctuating decline in precipitation from 1970 to 2020, hydrological processes and elements exhibited various changes, which in turn reflect the influence of temperature and precipitation on water resources. These changes indicate the temporal variability of the evolution of water resources in the Hanjiang River Basin.

The total water resources in the Hanjiang River Basin decreased by 8.92 million m^3 under the combined conditions (i.e., land use change and climate change), and the changes in water volume varied among sub-basins. Nevertheless, the changes in water volume caused by climate change were more consistent with the combined conditions, indicating that climate change is the main driver of the natural evolution of water resources in the Hanjiang River Basin.

The findings of this study may provide scientific guidance for determining the spatial and temporal distribution of water resources and hydrological processes at a basin scale in basins across the world with similar natural characteristics. The findings can also serve as a reference for investigating the spatial and temporal evolution of water resources in basins under changing conditions. The findings have potential applicability in integrated planning and efficient utilization of water resource in basins with similar conditions around the world, including southern China.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Materials, further inquiries can be directed to the corresponding author.

Author contributions

YL, JD, and CZ: Methodology, software, validation, formal analysis, investigation, resources, data curation, writing-original draft, writing-review and editing, visualization, supervision. YL, JD, MK, JZ, and CZ: Formal analysis, investigation, resources, data curation, writing-review and editing, visualization. YL, JD, MK, and CZ: conceptualization, methodology, software, formal analysis, investigation, resources, data curation, writing-original draft, writing-review and editing, visualization, supervision, project administration.

Funding

This research was supported by Forestry ecological monitoring network platform construction (Nos. 2021-KYXM-09; 2021-KYXM-09-001), and the Guangdong Natural Science Foundation (No. 2019A1515011627).

Acknowledgments

We thank the Field Scientific Observation and Research Station of Hanjiang River in Guangdong province, we also thank my other colleagues' valuable comments and suggestions that have helped improve the manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Abbaspour, K. C., Rouholahnejad, E., Vaghefi, S. R. I. N. I. V. A. S. A. N. B., Srinivasan, R., Yang, H., and Kløve, B. (2015). A continental-scale hydrology and water quality model for Europe: Calibration and uncertainty of a high-resolution large-scale SWAT model. *J. hydrology* 524, 733–752. doi:10.1016/j.jhydrol.2015.03.027

Abbaspour, K. C., Yang, J., Maximov, I., Siber, R., Bogner, K., Mieleitner, J., et al. (2007). Modelling hydrology and water quality in the pre-alpine/alpine Thur watershed using SWAT. *J. hydrology* 333 (2-4), 413–430. doi:10.1016/j.jhydrol.2006.09.014

Anand, J., Gosain, A. K., and Khosa, R. (2018). Prediction of land use changes based on Land Change Modeler and attribution of changes in the water balance of Ganga basin to land use change using the SWAT model. *Sci. total Environ.* 644, 503–519. doi:10.1016/j.scitotenv.2018.07.017

Arnold, J. G., Srinivasan, R., Muttiah, R. S., and Williams, J. R. (1998). Large area hydrologic modeling and assessment part I: Model development 1. *J. Am. Water Resour. Assoc.* 34 (1), 73–89. doi:10.1111/j.1752-1688.1998.tb05961.x

Brown, A. E., Zhang, L., McMahon, T. A., Western, A. W., and Vertessy, R. A. (2005). A review of paired catchment studies for determining changes in water yield resulting from alterations in vegetation. *J. hydrology* 310 (1-4), 28–61. doi:10.1016/j.jhydrol.2004.12.010

Davis, J., O'Grady, A. P., Dale, A., Arthington, A. H., Gell, P. A., Driver, P. D., et al. (2015). When trends intersect: The challenge of protecting freshwater ecosystems under multiple land use and hydrological intensification scenarios. *Sci. Total Environ.* 534, 65–78. doi:10.1016/j.scitotenv.2015.03.127

Feng, Y., and Hu, M. (2021). Spatial-temporal characteristics of extreme hydrological events in Hanjiang River basin. *Northwest Hydropower* 05, 1–9+16. doi:10.3969/j.issn.1006-2610.2021.05.001

Foley, J. A., DeFries, R., Asner, G. P., Barford, C., Bonan, G., Carpenter, S. R., et al. (2005). Global consequences of land use. *science* 309 (5734), 570–574. doi:10.1126/science.1111772

Gassman, P. W., Reyes, M. R., Green, C. H., and Arnold, J. G. (2007). The soil and water assessment tool: Historical development, applications, and future research directions. *Trans. ASABE* 50 (4), 1211–1250. doi:10.13031/2013.23637

Hagemann, S., Chen, C., Clark, D. B., Folwell, S., Gosling, S. N., Haddeland, I., et al. (2013). Climate change impact on available water resources obtained using multiple global climate and hydrology models. *Earth Syst. Dyn.* 4 (1), 129–144. doi:10.5194/esd-4-129-2013

Hall, J. W., Grey, D., Garrick, D., Fung, F., Brown, C., Dadson, S. J., et al. (2014). Coping with the curse of freshwater variability. *Science* 346 (6208), 429–430. doi:10.1126/science.1257890

Krysanova, V., and White, M. (2015). Advances in water resources assessment with SWAT—an overview. *Hydrological Sci. J.* 60 (5), 771–783. doi:10.1080/02626667.2015.1029482

Kundu, S., Khare, D., and Mondal, A. (2017). Individual and combined impacts of future climate and land use changes on the water balance. *Ecol. Eng.* 105, 42–57. doi:10.1016/j.ecoleng.2017.04.061

Kvålseth, T. O. (1985). Cautionary note about R 2. *Am. Statistician* 39 (4), 279–285. doi:10.1080/00031305.1985.10479448

Legates, D. R., and McCabe, G. J., Jr (1999). Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. *Water Resour. Res.* 35 (1), 233–241. doi:10.1029/1998WR900018

Li, C., and Fang, H. (2021). Assessment of climate change impacts on the streamflow for the mun River in the mekong basin, southeast Asia: Using SWAT model. *Catena* 201, 105199. doi:10.1016/j.catena.2021.105199

Li, Z., Huang, B., Qiu, J., Cai, Y., Yang, Z., and Chen, S. (2021). Analysis on evolution characteristics of ecological flow of Hanjiang River under changing environment. *Water Resour. Prot.* 37 (05), 22–29. doi:10.3880/j.issn.1004-6933.2021.05.004

Ljungqvist, F. C., Krusic, P. J., Sundqvist, H. S., Zorita, E., Brattström, G., and Frank, D. (2016). Northern Hemisphere hydroclimate variability over the past twelve centuries. *Nature* 532 (7597), 94–98. doi:10.1038/nature17418

Long, H., Lin, B., Ou, Y., and Chen, Q. (2019). Spatio-temporal analysis of driving factors of water resources consumption in China. *Sci. Total Environ.* 690, 1321–1330. doi:10.1016/j.scitotenv.2019.06.311

Lotz, T., Opp, C., and He, X. (2018). Factors of runoff generation in the Dongting Lake basin based on a SWAT model and implications of recent land cover change. *Quat. Int.* 475, 54–62. doi:10.1016/j.quaint.2017.03.057

McCuen, R. H., Knight, Z., and Cutter, A. G. (2006). Evaluation of the nash–sutcliffe efficiency index. *J. Hydrol. Eng.* 11 (6), 597–602. doi:10.1061/(ASCE)1084-0699(2006)11:6(597)

Miller, J. D., Kim, H., Kjeldsen, T. R., Packman, J., Grebby, S., and Dearden, R. (2014). Assessing the impact of urbanization on storm runoff in a peri-urban catchment using historical change in impervious cover. *J. Hydrology* 515, 59–70. doi:10.1016/j.jhydrol.2014.04.011

Nash, J. E., and Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—a discussion of principles. *J. hydrology* 10 (3), 282–290. doi:10.1016/0022-1694(70)90255-6

Onyutha, C. (2022). A hydrological model skill score and revised R-squared. *Hydrology Res.* 53 (1), 51–64. doi:10.2166/nh.2021.071

Osei, M. A., Amekudzi, L. K., Wemegah, D. D., Preko, K., Gyawu, E. S., and Obiri-Danso, K. (2019). The impact of climate and land-use changes on the hydrological processes of Owabi catchment from SWAT analysis. *J. Hydrology Regional Stud.* 25, 100620. doi:10.1016/j.ejrh.2019.100620

Piao, S., Ciais, P., Huang, Y., Shen, Z., Peng, S., Li, J., et al. (2010). The impacts of climate change on water resources and agriculture in China. *Nature* 467 (7311), 43–51. doi:10.1038/nature09364

Ross, E. R., and Randhir, T. O. (2022). Effects of climate and land use changes on water quantity and quality of coastal watersheds of Narragansett Bay. *Sci. total Environ.* 807, 151082. doi:10.1016/j.scitotenv.2021.151082

Sajikumar, N., and Remya, R. S. (2015). Impact of land cover and land use change on runoff characteristics. *J. Environ. Manag.* 161, 460–468. doi:10.1016/j.jenvman.2014.12.041

Santhi, C., Arnold, J., Williams, J. R., Dugas, W. A., Srinivasan, R., and Hauck, L. M. (2001). Validation of the SWAT model on a large river basin with point and nonpoint sources. *J. Am. Water Resour. Assoc.* 37, 1169–1188. doi:10.1111/j.1752-1688.2001.tb03630.x

Shrestha, S., Bhatta, B., Shrestha, M., and Shrestha, P. K. (2018). Integrated assessment of the climate and landuse change impact on hydrology and water quality in the Songkhram River Basin, Thailand. *Sci. Total Environ.* 643, 1610–1622. doi:10.1016/j.scitotenv.2018.06.306

Sivakumar, B. (2011). Global climate change and its impacts on water resources planning and management: assessment and challenges. *Stoch. Environ. Res. Risk Assess.* 25 (4), 583–600. doi:10.1007/s00477-010-0423-y

Souffront Alcantara, M. A., Nelson, E. J., Shakya, K., Edwards, C., Roberts, W., Krewson, C., et al. (2019). Hydrologic modeling as a service (HMaaS): a new approach to address hydroinformatic challenges in developing countries. *Front. Environ. Sci.* 7, 158. doi:10.3389/fenvs.2019.00158

Sterling, S. M., Ducharne, A., and Polcher, J. (2013). The impact of global land-cover change on the terrestrial water cycle. *Nat. Clim. Chang.* 3 (4), 385–390. doi:10.1038/nclimate1690

Strohbach, M. W., Döring, A. O., Möck, M., Sedrez, M., Mumm, O., Schneider, A. K., et al. (2019). The "hidden urbanization": Trends of impervious surface in low-density housing developments and resulting impacts on the water balance. *Front. Environ. Sci.* 7, 29. doi:10.3389/fenvs.2019.00029

Uniyal, B., Jha, M. K., and Verma, A. K. (2015). Assessing climate change impact on water balance components of a river basin using SWAT model. *Water Resour. manage.* 29 (13), 4767–4785. doi:10.1007/s11269-015-1089-5

Wang, G. Q., Zhang, J. Y., Xuan, Y. Q., Liu, J. F., Jin, J. L., Bao, Z. X., et al. (2013). Simulating the impact of climate change on runoff in a typical river catchment of the Loess Plateau, China. *J. Hydrometeorol.* 14 (5), 1553–1561. doi:10.1175/JHM-D-12-081.1

Wu, S., Gao, J., Wei, B., Zhang, J., Guo, G., Wang, J., et al. (2021). Theoretical paradigm for natural disaster-resilient society. *Acta Geogr. Sin.* 76 (5), 1136–1147. doi:10.11821/dlxb202105007

Wu, Y., Liu, S., and Gallant, A. L. (2012). Predicting impacts of increased CO2 and climate change on the water cycle and water quality in the semiarid James River Basin of the Midwestern USA. *Sci. Total Environ.* 430, 150–160. doi:10.1016/j.scitotenv.2012.04.058

Xia, J., Shi, W., Luo, X., Hong, S., Ning, L., and Gippel, C. J. (2015). Revisions on water resources vulnerability and adaption measures under climate change. *Adv. Water Sci.* 26 (2), 279–286. doi:10.14042/j.cnki.32.1309.2015.02.019

Zang, C. F., Liu, J., Van Der Velde, M., and Kraxner, F. (2012). Assessment of spatial and temporal patterns of green and blue water flows under natural conditions in inland river basins in Northwest China. *Hydrol. Earth Syst. Sci.* 16 (8), 2859–2870. doi:10.5194/hess-16-2859-2012

Zang, C., Liu, J., Gerten, D., and Jiang, L. (2015). Influence of human activities and climate variability on green and blue water provision in the Heihe River Basin, NW China. *J. Water Clim. Chang.* 6 (4), 800–815. doi:10.2166/wcc.2015.194

Zang, C., and Liu, J. (2013). Trend analysis for the flows of green and blue water in the Heihe River basin, northwestern China. *J. Hydrology* 502, 27–36. doi:10.1016/j.jhydrol.2013.08.022

Zhang, J., Wang, G., Jin, J., He, R., and Liu, C. (2020a). Evolution and variation characteristics of the recorded runoff for the major rivers in China during 1956—2018. *Adv. Water Sci.* 31 (2), 153–161. doi:10.14042/j.cnki.32.1309.2020.02.001

Zhang, L., Cheng, L., Chiew, F., and Fu, B. (2018). Understanding the impacts of climate and landuse change on water yield. *Curr. Opin. Environ. Sustain.* 33, 167–174. doi:10.1016/j.cosust.2018.04.017

Zhang, Q., Liu, Y., Yao, J., Lai, X., Li, X., Wu, G., et al. (2020b). Lake hydrology in China: Advances and prospects. *J. Lake Sci.* 32, 1360–1379. doi:10.18307/2020.0511

Zhang, Z. (2007). *Synthetic study of land use/cover change in typical area in Hanjiang River basin, Guangdong province. PhD*. Guangzhou: Graduate University of Chinese Academy of Sciences. Guangzhou Institute of Geochemistry. CNKI: CDMD:1.2007.101649.

Zhou, C., Xie, Y., Zhang, A., Liu, C., and Yang, J. (2022). Spatiotemporal analysis of interactions between seasonal water, climate, land use, policy, and socioeconomic changes: Hulun-Buir Steppe as a Case Study. *Water Res.* 209, 117937. doi:10.1016/j.watres.2021.117937

Zhou, G., Wei, X., Chen, X., Zhou, P., Liu, X., Xiao, Y., et al. (2015). Global pattern for the effect of climate and land cover on water yield. *Nat. Commun.* 6 (1), 5918–5919. doi:10.1038/ncomms6918

frontiers | Frontiers in Environmental Science

TYPE Original Research
PUBLISHED 05 January 2023
DOI 10.3389/fenvs.2022.857082

OPEN ACCESS

EDITED BY
Fei Tian,
China Agricultural University, China

REVIEWED BY
Subash Surendran Padmaja,
National Institute for Agricultural Economics and Policy Research (NIAP), India
Keshav Lall Maharjan,
Hiroshima University, Japan

*CORRESPONDENCE
Hengyun Ma,
h.y.ma@163.com

SPECIALTY SECTION
This article was submitted to Environmental Economics and Management, a section of the journal Frontiers in Environmental Science

RECEIVED 18 January 2022
ACCEPTED 18 November 2022
PUBLISHED 05 January 2023

CITATION
Ahmad MI, Oxley L, Ma H and Liu R (2023), Does rural livelihood change? Household capital, climate shocks and farm entry-exit decisions in rural Pakistan.
Front. Environ. Sci. 10:857082.
doi: 10.3389/fenvs.2022.857082

Does rural livelihood change? Household capital, climate shocks and farm entry-exit decisions in rural Pakistan

Muhammad Irshad Ahmad[1], Les Oxley[2], Hengyun Ma[1]* and Ruifeng Liu[1]

[1]College of Economics and Management, Henan Agricultural University, Zhengzhou, Henan, China, [2]School of Accounting, Economics and Finance, University of Waikato, Hamilton, Waikato, New Zealand

Agriculture is a primary source of livelihoods in developing countries. The process of entry and exit of farming activities continues to play an important role in maintaining competition in agriculture and allocating resources between agriculture and other sectors. However, climate changes and other economic and social shocks have been severely affecting farmers' livelihoods. This article examines rural household livelihood transition in the context of farm entry and exit decisions in rural Pakistan. Using 1867 rural household survey data, we brought insights into how livelihood assets, climate shocks, climate investment and locational characteristics affect their farm entry and exit decisions. The results indicate that the proportion of farm entry (24%) was higher than that of farm exit (15%). The major factors were crop inputs using as credit with a huge markup, crop inputs sold by farmers on net cash in financial crisis, climate shocks and poor climate investment that contributed to farm exit. They were household head characteristics, land ownership (family farm), and livestock ownership that increased the likelihood of farm entry decisions. Farm exit decisions were significantly and positively associated with household migration status, irrigation water shortage, off-farm income, crop input used as credit, crop diseases, climate shocks and lack of local government role in sharing climate knowledge. Furthermore, Propensity Score Matching (PSM) results found that the entry decision significantly decreases household income, while the exit decision significantly increases household income and food security status. These findings provide insights into farm entry and exit for those who are planning livelihood transition, and offer recommendations on how to overcome the constraints faced by farming businesses, agricultural sustainability, self-sufficiency and food security during the transition nationally and internationally.

KEYWORDS

livelihood assets, farm entry, farm exit, livelihoods transition, off-farm work, on-farm work, farmer wellbeing

1 Introduction

Farm entry and exit are a process of livelihood diversification or transition in the agricultural sector, which contributes to global competitiveness of agriculture and efficient resources allocation between agriculture and other sectors in the economy (Binswanger-Mkhize, 2013). More than 2.5 billion out of the three billion rural population deriving their livelihoods from agriculture (FAO, 2021). However, the dwindling size of agricultural land, increasing population, low productivity and hostile agro-ecological factors often result in extreme income variability in agriculture. In response, rural households often use farm entry and exit to diversify their livelihood activities and smooth income variability. Off-farm livelihood opportunities in rural areas often play an important role in reducing food insecurity (Barrett et al., 2001). However, farm exit in the form of shifting from on-farm to off-farm activities does not necessarily create positive outcomes in terms of either reducing poverty or increasing incomes, particularly in developing agrarian countries. In some circumstances, this shift seems to have increased poverty (Imai et al., 2014).

Theoretically, structural transformation in the economy is driving on-farm labor into the off-farm sector for pursing a more sustainable livelihood. Some farming households exit farm and move to other sectors, as they consider agriculture a low productive and highly risky occupation (Haggblade et al., 2010; Hussain, 2014). Farm exit to off-farm sectors has been extensively investigated (Kimhi and Bollman, 1999; Pietola et al., 2002; Glauben et al., 2006; World Bank, 2007; Cai and Wang, 2010; Knight et al., 2011; Wang et al., 2011; Bhandari, 2013; Ahmad et al., 2020). In contrast, the shift from off-farm to on-farm as new farm entry has not been investigated yet in the context of agricultural-based economies including Pakistan. In such cases, promoting agricultural activities remains a priority in order to achieve the Millennium Development Goal (MDG), which is to reduce poverty and hunger, to sustain livelihoods by increasing new and well educated farmers and by stop farm exits as well (World Bank, 2008). Adjustments to an agricultural structure could also attract and encourage more people to enter farming and pursue farming either as a main occupation or an additional source of income (Mishra and El-Osta, 2016). In addition, the emphasis of farmers' rights for self-sufficiency in food by growing their own food can encourage former farmers to return to farming (Agarwal and Agrawal, 2017).

Pakistan is heavily dependent on agricultural production which contributes 24% to its GDP (Pakistan Bureau of Statistics, 2019). Unfortunately, its agricultural system fails to maintain its growth due to serious challenges such as water shortage, climatic change, rising input prices, limited policy incentives for farming and low trust in government, making farmers reduce their cultivated area and worsen the country's overall agricultural productivity (Pakistan Bureau of Statistics, 2018). Following the 9/11 event in U.S. in 2001, Pakistan has fought a long war of 19 years against terrorism as an ally of the US, by closing doors to foreign investments. Since then, the farm sector has been in its recession. Pakistan's agriculture sector started to decrease in sizes since 2001 due to emerging water shortages, climatic changes and natural disastrous events (floods, heavy rain and drought), as well as high input prices, low output prices, which reduced the earnings of both farming and non-farming communities, and increased the unemployment rate in the country (State Bank of Pakistan, 2015). More importantly, the challenges mentioned above are causing farmers to exit from farming. For example, the agricultural sector has experienced an obvious decline at both the absolute and relative levels in farm employment percentage, from 45 to 38.5% over the last decade (Pakistan Bureau of Statistics, 2018). It is thus of importance to know why farmers exit agriculture and how their livelihoods change after exit (Ahmad et al., 2020).

This study aims to firstly close a significant research gap by identifying the factors driving farm entry and exit in Pakistan. Secondly, this study further investigates the impact of farm entry and exit on households' wellbeing in terms of total household income, food security status and ability to save for an emergency funds. Thirdly, this study also aims to identify the beginner farmers, who are they, and why did they enter into farming? To the best of our knowledge, these can be the first empirical work to examine both farm entry and exit decisions for rural households in Pakistan.

To achieve the goals above, this study is organized as follows: The next section provides a comprehensive literature review, followed by introducing farmers' livelihood options in Pakistan. Section 4 provides a conceptual framework, followed by introducing our methods and data. Section 6 provides estimated results and analyses, followed by a balance test and sensitive analyses. The last section concludes.

2 Literature review

Over the past century, agriculture sector has transformed from a labor-intensive to a capital-intensive industry. The shift has allowed people to engage in secondary and tertiary sectors and to relocate to non-farm regions. As a result, farming workers have fallen significantly (Lobao and Meyer, 2001; Gale 2003; Conkin, 2008; Ahearn and Newton, 2009). For example, Gale (2003) noted that most farm exits are voluntary, retiring, passing management to the next-generation or leaving farming due to poor health and death in western countries. The family farm is viewed as the backbone of rural communities and the decline of farm number raises questions with regard to whether these communities can sustain themselves. Although intergenerational family transfer remains the dominant mechanism for farm succession, in most western European countries and the United States, the number of family farm transfers was decreasing (Gale, 2003; Calus et al., 2008).

For the new entry farm, there could be some barriers for the new farmers. For example, the studies consistently identify access to affordable land as the greatest barrier to entry to agriculture (Ackoff et al., 2017; Frost, 2017). Many beginning farmers said that they lacked the means to employ the number of skilled farm workers necessary to maintain and grow farm operations (Ackoff et al., 2017). Climate change could be the key issue causing an increase in unpredictable extreme weather events; increases in drought and flooding events threaten to destroy crops and reduce yields and most of the farmers have experienced the influence of climate changes (Ackoff et al., 2017). Therefore, many small- and mid-size farm operators require off-farm income to make ends meet (Gillespie and Johnson, 2010). In general, beginning farmers face steep start-up costs and barriers to accessing capital, land, and credit (Ahearn, 2011; Lusher Shute, 2011; Calo, 2018).

Farming is characterized by an ageing population with a reduced rate of entry into farming by younger farmers and a reduced rate of retirement by older farmers (ADAS, 2004). The farming industry has failed to attract "new blood" into the industry, partly due to the poor rewards and partly due to entry barriers such as high start-up costs and a shortage of available land (ADAS, 2004); something that is exacerbated by restructuring processes that are leading to fewer, larger farms within both the private and county estates (Whitehead and Millard, 2000). At the same time, a lack of suitable successors and taxation issues have been identified as making farmers reluctant to retire (Williams and Farrington, 2006). This is the result of a number of entry-exit challenges such as increasing capital requirements, low expected rates of return and higher off-farm career opportunities (Gale, 2003; Williams and Farrington, 2006). Several studies from developed countries have highlighted these adjustment challenges facing the farming industry (Caskie et al., 2002; Errington and Lobley, 2002; ADAS, 2004; Calus et al., 2008).

Concerns about the sustainability of an ageing farming population have brought interest in so called entry-exit issues in policy circles. Policy interventions to date have offered limited scope in stimulating farm transfer, however, the increase in unconventional tenures which include partnerships, share farming and contract farming, would appear to offer new opportunities for those wishing to enter or leave farming (Ingram and Kirwan, 2011). Bruce (2019) identified a new pathway into alternative agriculture that returning farmers come from farm families, but left agriculture to pursue higher education or a non-farm career and then re-entered agriculture later in life through Alternative Food Networks (AFNs). However, social movements promoting alternative models of agriculture have created organizations to support a new generation of farmers, and generated AFNs that provide new training opportunities and markets for aspiring farmers.

3 Farmers' livelihood options in Pakistan

As most of agriculture farms in Pakistan are small and not well educated, thus they do not have better off-farm jobs in the country except daily paid labor. For example, a study from Khyber Pakhtunkhwa (KPK) province investigated the determinants of the off-farm employment of the small farm which showed that most of small farms (90%) were engaged in off-farm jobs along with agricultural activities (Ali et al., 2014). The nature of their job was in daily paid labor, part time employment, and different off-farm businesses. The effects of farm underemployment, working age group size (age of the farmer), income from other sources, and education were positive on the off-farm employment (Ali et al., 2014). For example, Rizwan et al. (2017) conducted a study in province Punjab, Pakistan and found that about 66% farmers were involved with off-farm activities along with on-farm activities. The results indicated that education has significant influence and stimulate for engagement in off-farm employment. However, presence of younger population in households and land renting opportunity stimulate migration in other cities and countries. Dependency ratio and large family size were the driving factors for participation in off-farm labour activities.

Though, off-farm activities as part-time are also being performed in Gilgit-Baltistan, Pakistan. In this case, farmer characteristics (e.g. farmer age, gender and education), farm characteristics (e.g., farm size, specialization in horticulture, etc.) and agricultural income (Shahzad et al., 2021). Tahir et al. (2012) investigated the factors contributing to off-farm employment in North West Pakistan. They found that farm size, family size, farm underemployment, education, and income from other sources were the main factors determining off-farm employment. It was also observed that farmers of the comparatively developed areas devote more time to off-farm employment. The study revealed that most of the farmers were engaged in daily paid labor. Overall there is a gradual shift from farm to off-farm employment which is resisted by the underdeveloped means of transport and communication, education and lack of basic infrastructure.

Few studies that investigated the factors that affect occupational choices of populations living in rural areas of Pakistan. Jan et al. (2012) revealed that the likelihood to participate in non-farm informal sector increases for household having relatively younger head with no education. Household size positively and significantly related to all the occupational groups while additional working members in a household reduces the odds to engage in farming by about 67% relative to non-farm informal sector. Similarly, *per capita* income also plays a significant role in pursuing occupations other than informal activities. In addition, to know the influence of migration on farm exit, Abbasi and Kim (2018) investigated that agriculture is not the primary source of income and 32%

agriculture labor force migrated and migration is the main reason for declining labor force and increasing shifted to off farm activities.

However, the youth (aged between 15 and 29 years) in Pakistan seems less interested in performing agriculture activities. They regarded the agriculture sector as non-profitable, hence they do not see them joining the agriculture sector due to high cost of production, crashed marketing system, absence of farmer-friendly policies, environmental issues and lack of support from government in agriculture sector (Ahmad et al., 2020; Aftab et al., 2021). Thus, most of farms are involved in non-farm income generation activities in southern Punjab, Pakistan. The majority of the farmers offered labor for off-farm work followed by self-employment ventures. The major reason to pursue non-farm work includes low income from agriculture, mitigating risks associated with farming. A range of socioeconomic and infrastructure-related variables are associated with the decision to participate in specific off-farm activity, such as age, education, family size, farm income, dependency burden, farming experience, and distance to the main city.

In case of Pakistan, there has been a steady shift from subsistence farming to cash crops and fruit production, which is particularly noticeable in the accessible parts of the region that are located closer to urban centers. In addition, the rising proportion of household income from non-farm activities (increasing from 43% in 1994 to 63% in 2005 and more than 70% in 2020 (Shahzad et al., 2021), plays an important role in the transformation of the rural economy. Similarly, the increased labor outmigration towards the services sector in down-country Pakistan stemming from the improved formal education systems has increased the share of non-farm employment plays an important role in the transformation of the rural economy. Similarly, the increased labor outmigration towards the services sector in down-country Pakistan stemming from the improved formal education systems has increased the share of non-farm employment.

Under these circumstances, the maintenance of farming communities is largely under threat. The increased rate of rural-to-urban migration (Gioli et al., 2014), particularly that of younger people (Benz, 2016) and the rapid growth of the non-farm sector (Gioli et al., 2014; Shahzad et al., 2021) have resulted in decreased agricultural land-use and increased uncertainty regarding farm continuation.

4 Conceptual framework

Studies on rural off-farm and on-farm activities consider livelihood diversification and income stabilization (or risk minimization) as the major motives for working outside of agriculture (Rose, 2001; Haggblade et al., 2010). Livelihood diversification is driven by "pull" factors (e.g., markets, opportunities, infrastructural facilities and supportive institutions) and "push" factors (e.g., various idiosyncratic shocks such as floods, droughts, environmental degradation, chronic rainfall deficit). Similarly, off-farm labor could shift towards farming activities in case of unemployment, old age, job insecurity and health issues (Mishra and El-Osta, 2016). Also, farmers' rights for self-sufficiency in food by growing their own food can encourage former farmers to return to farming (Agarwal and Agrawal, 2017). As a result, a large part of the world's labor force work in agriculture, not by choice, but due to lack of alternatives (Cain, 1977; Kumar and Hotchkiss, 1988; Chitrakar, 1990; Karan and Ishii, 1995; Filmer and Pritchett, 1997; Agarwal, 2014). It is important to understand households' motives of diversifying beyond agriculture or moving away from off-farm work. Therefore, we base our theoretical framework on livelihood vulnerability as diversification/transition motive, because it best serves the main objective of the study, which is to investigate the determinants of households' diversification decisions[1]of farm entry and exit.

To better understand the Pakistani agriculture sector, we classify the factors associated with "farm entry and exit" into the following groups: 1) human capital referring to characteristics of household head and household; 2) natural capital including land, livestock and irrigation systems; 3) economic capital comprising of loans or credit, off-farm employment, and off-farm income and sources; 4) climate shocks, including natural disasters and severe crop diseases; 5) climate change investment[2] referring to access to micro finance institutions (MFIs), all weather road and climate knowledge (Eifert and Ramachandran, 2004); and 6) locational characteristics such as home remoteness and the extent of commercialization and urbanization.

It is hypothesized that a household can involve in one of the two livelihood transition scenarios: 1) exit farming and shift to off-farm activities; or 2) enter farming and shift to on-farm activities. Each household is assumed to make a rational choice, which is when income generated from the new sector is higher than that from the last. Households that intends to enter farming face barrier such as capital investment, climate challenges, farming experience, knowledge and skills and other household constraints. Similarly, households that want to exit farming could face barriers in off-farm employment, age, skilled labor, education, family labor and other household constraints. We assume that a farming household i has fixed capital and labor

1 Even if the motive of the household is higher income, climate shocks affecting agricultural income may still influence its expected earnings, and hence the diversification decision.

2 Climate change investment is defined as different characteristics specific to a certain location that could act as incentives or disincentives for entry or exit such as availability/unavailability of financial services, infrastructure, governance and regulations etc. (Eifert and Ramachandran, 2004).

endowments, assigned among different activities in agriculture, which is expressed in Equation One:

$$I_i = E_t \sum_{\tau=1}^{T} \beta^{\tau-1} \pi_i \left(IP, OP, LC_i, \varphi_t, \varepsilon_i\right) \tag{1}$$

Where I_i is the income of household i, E_t is expectation operator providing information at time t, β is the subjective discount factor, T is the number of periods, π_i isthe profit generated for household i which is a function of input price IP, output price OP, fixed labor and capital endowments LC_i, a vector of economic shocks that could affect household income and livelihood diversification φ_t, and ε_i includes unobserved characteristics that could affect income. Suppose a household is generating income from agricultural activities, Equation One is adapted to Equation Two with subscript A indicating agriculture:

$$I_{Ai} = E_t \sum_{\tau=1}^{T} \beta^{\tau-1} \pi_{A,i} \left(IP_A, OP_A, LC_i, \varphi_t, \varepsilon_i\right) \tag{2}$$

For farm entry, household income is denoted by (I_{B_i}) and expressed in Equation Three:

$$I_{Bi} = -C_{i,t}\left(N_i,\ I_i,\ H_i,\right) + E_t \sum_{\tau=1}^{T} \beta^{\tau-1} \pi_{Bi}$$

$$\left(IP_B,\ (N_i,\ I_i), OP_B\left(N_i,\ I_i\right), LC_{B,i}, \mu_i\right) \tag{3}$$

Where $C_{i,t}$ is farm entry cost, which could be affected by investment I_i such as climate, capital and farm machinery assets and other inputs. Furthermore, other factors like availability of financial services, government policies, taxes and infrastructure could reduce or increase the entrance barrier. These characteristics could also capture input-output markets at specific locations and availability of micro finance institutions (MFIs) and farm advisory services. Furthermore, farm entry could also be affected by the characteristics of household head (H_i) such as age, education, migration status and location (N_i). Input prices (IP_B) and output prices (OP_B) are part of the entry function into farming, which could be affected by location factors that may be associated with lower than market prices for crop outputs[3] and μ_i isunobserved characteristics.

Households assign their total amount of fixed labor and capital for different activities, which can be expressed as:

$$LC_i = LC_{A_i} + LC_{B_i} \tag{4}$$

As we mentioned above, household could diversify income strategy through farm entry if on-farm income is greater than off-farm income, as shown in Eq. 6.

$$I_{A_i} + I_{B_i} > I_i \tag{5}$$

OR

$$I_{B_i} > I_i - I_{A_i} \tag{6}$$

Following Eq. 6, the probability of diversifying to farming can be written as:

$$prob_{(B_i)} = prob\left(-C_{i_t}\left(N_i,\ I_i,\ H_i,\right) + E_t \sum_{\tau=1}^{T} \beta^{\tau-1} \pi_{Bi}\left(IP_B, (N_i,\ I_i),\right.\right.$$

$$OP_B\left(N_i,\ I_i\right), LC_{B_i}, \mu_i)) > prob\left(E_t \sum_{\tau=1}^{T} \beta^{\tau-1} \pi_i\right.$$

$$\left.\left.\left(IP, OP, LC_i, \varphi_t, \varepsilon_i\right) - \pi_{Ai}\left(IP_A, OP_B, LC_i, \varphi_t, \varepsilon_i\right)\right)\right) \tag{7}$$

As input and output prices do not change whether the household works only in agriculture or diversity to off-farm work. Therefore, we do not expect them to play a significant role in affecting household's choice except when the actual level of profit is estimated. Hence, we expect that household labor and capital endowments, which are fixed, may play a vital role in decisions on farm entry and exit.

If a household enters into farming, d_{B_i} is defined as 1, and 0 if a household stays in off-farm activities (see Supplementary Appendix SA1). Stochastic factors ε_i and μ_i are assumed to be identically and independently distributed, and then the probability of farm entry is given as:

$$\begin{cases} f\left(H_i, N_i, I_i, \varphi_t\right) & if\ d_{B_i} = 1 \\ 0 & otherwise \end{cases} \tag{8}$$

Further, once households enter into farming, they could face two choices: either continue or exit farming. Households could face barriers when existing farming, such as investment made on non-transferable fixed assets, land rent, farm machinery and long run investment. Whereas, incentives at household level such as high grain prices could encourage farm entry, particularly for those who enter to produce domestic grain food. The income generated from agriculture will then depend on the trade-off between the cost of farm exit and the profit earned from continuing farming. Moreover, households' earnings depend upon farming inputs and output prices which are determinants of farm entry together with other factors such as land status, yields, etc. Given these, if the present value of on-farm income is less than off-farm income, household will decide to exit farming, which can be expressed as:

$$I_{B_i}^{E} < I_i^{E} - I_{A_i}^{E} \tag{9}$$

where I_i^E is income of an incumbent household i. Likewise, the probability of farm exit is a function of household specific characteristics H_i, farming capital factors FC_i, fixed inputs of labor and capital LC_i, locational specific factors (N_i, I_i) and climate shocks affecting agriculture (φ_i).

3 Firstly, most of the farms in Pakistan are small and did not have access to commercial markets. Therefore, they are dealt by commission agents or middle men and they are offered a low price for their output. Secondly, farmers who use crop inputs on credit incur a heavy markup and are bounded to sell their crop outputs to input dealers or lenders at a low price.

TABLE 1 Study provinces and sample size.

Province	Number of districts	Exit farming	Continue farming	Stay off-farm	Enter into farming	Total sample
Punjab	12	62	460	533	131	1,186
Sindh	5	60	216	136	60	472
KPK	2	14	113	47	35	209
Total	**19**	**136**	**789**	**716**	**226**	**1867**

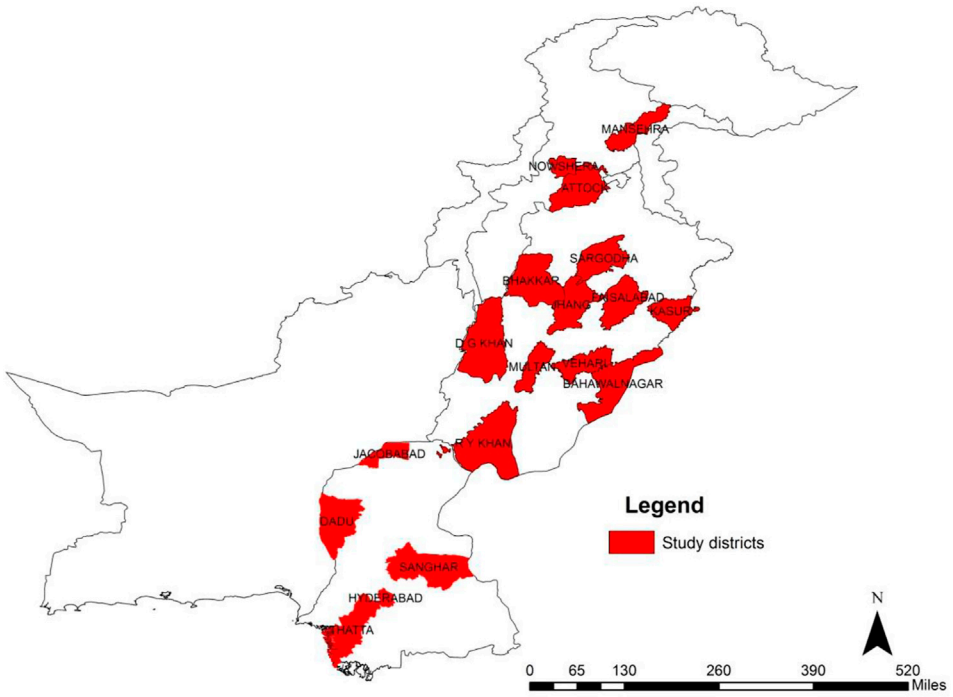


FIGURE 1
Study districts and map of Pakistan.

5 Methods and data

5.1 The data

Data for this study is derived from the Pakistan Rural Household Panel survey conducted in the Punjab, Sindh and Khyber-Pakhtunkhwa (KPK) provinces of Pakistan in 2012–2014.

The survey was designed and supervised by International Food Policy Research Institute (IFPRI) and was administered by Innovative Development Strategies (IDS), Islamabad, Pakistan. IDS served as the data collector and handled all of the survey logistics, from enumerator training to the processing of the completed questionnaires. This panel survey contains three different rounds: Rounds 1, Round 1.5, and Round 2, which identifies the status of household either continue or exit farming and their climate change adaptation measure at farm level (International Food Policy Research Institute, 2014; International Food Policy Research Institute, 2015; International Food Policy Research Institute, 2016; International Food Policy Research Institute, 2017). Additionally, one of author personally took part in data collection when surveys were conducted in 2012 and 2014 in all three rounds.

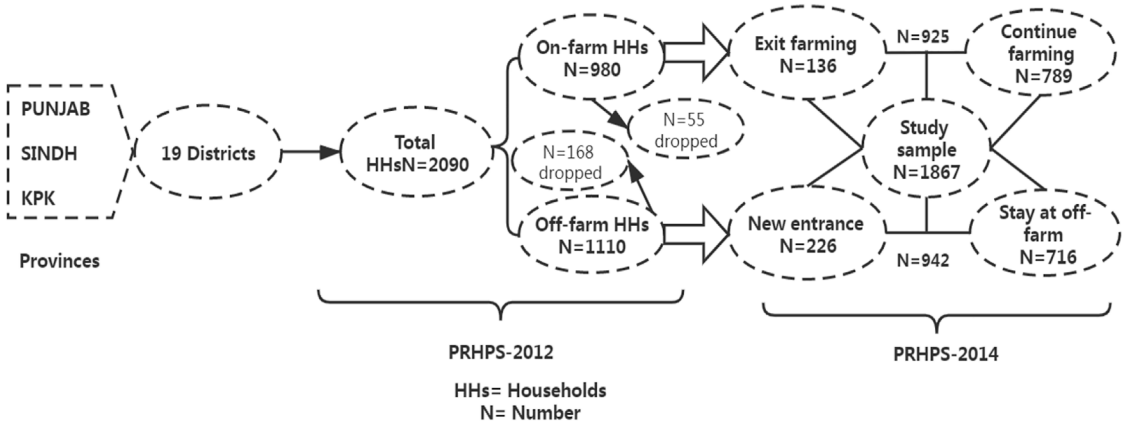


FIGURE 2
Sampling of Households (HHs) and study process from the Pakistan Rural Household Panel Survey (PRHPS, 2012–2014).

A total of 19 districts were surveyed across three provinces: 12 from Punjab, 5 from Sindh and 2 from KPK (Table 1; Figures 1,2[4]). Within each district, 4 mouzas[5] were chosen as Primary Sampling Units (PSU) using an equal probability systematic selection approach. The lists of revenue villages/mouzas/dehs provided by the 1998 Population Census were used as the sampling frame. The enumeration teams sectioned each mouza into enumeration blocks according to the village map. Each block consists of a maximum of 200 households. Subsequently, one enumeration block was randomly chosen from each mouza and households within a PSU were considered as Secondary Sampling Units (SSU). Then 28 households were randomly selected from each block of a maximum of 200 households. Finally, households (HHs) were defined as, 'a family or group of persons living in common accommodation (family members living in the same building or boundary wall), and cooking or sharing all their meals together'. The respondents were the most knowledgeable member and major decision maker in domestic affairs within the household.

The survey includes various types of information on human capital, natural capital, economic capital, locational characteristics, as well as household demographic and socioeconomic characteristics. We utilize the surveyed household panel dataset (2012–2014) to identify farm entry and exit of rural households as a whole. The first round survey in 2012 includes additional household information for the year 2010 and 2011. For example, the 2012 survey contains information on employment of the households such as whether they were working on-farm[6] (farmer) or off-farm[7] (not a farmer) in year 2010 and 2011 as well as in 2012, respectively. The 2014 survey asked the same question for 2012, 2013 and 2014.

Out of the total 2090 surveyed households, 1,110 (53%) were working off-farm, while the remaining 980 (47%) were working on-farm in 2012. We then matched these1110[8] households with those in the 2014 survey to identify any "new entrant"[9] into farming. It is found that 226 (24%) off-farm households in 2012 entered farming in 2014 and were defined as "new entrants" (see Table 1; Figures 2, 3). We matched 980[10] households with the 2014 survey to identify any "farm exit". It is found that 136 (15%) households exited farming in 2014, and were defined as "exit farming" (Figure 4).

4 Baluchistan province was not surveyed and skipped due to security reasons. The sample excludes rural areas in Baluchistan and the Federally Administered Tribal Areas because they were considered unsafe for the enumeration.

5 In Pakistan, province subunit is district, then district subunit is Tehsil, whereas Tehsil subunit is Union Council, and Union Council subunit is Mouza, and Mouza subunit is village/dehs, village/dehs subunit is Basti/Chak (groups of several households/families lived in and are identified by their Basti/Chak).

6 Household head/individual(s) from a family who cultivated farmland, despite the fact that any of family member worked at off–farm are considered as on–farm households.

7 Household head individual(s) from a household/family who did not cultivate farmland since 2010–2014 are considered as off–farm households.

8 From 1,110 households, 168 observations were dropped due to missing data and therefore 942 households remained in 2014.

9 "New entrants" into farming referred to households who did not cultivate land and were working off–farm since 2010 but started agricultural activities for the first time in 2014. Moreover, inheriting a family farm is also considered as "new entrant".

10 55 observations were dropped due to missing data and therefore 925 households remained in 2014.

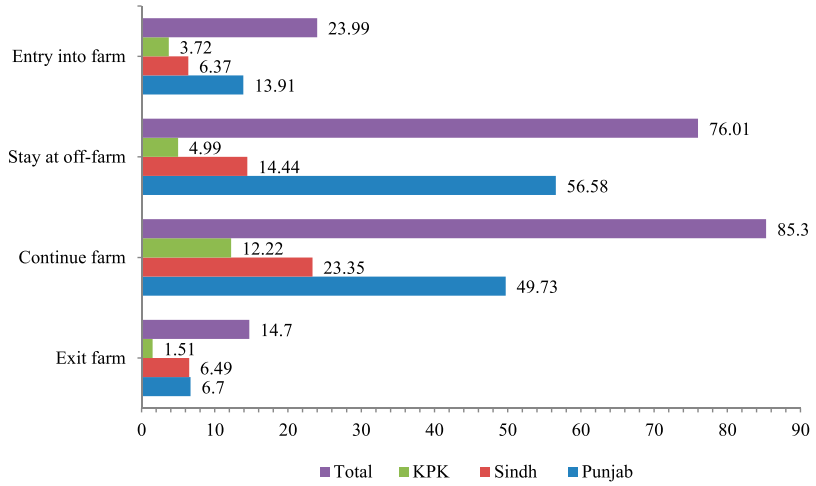


FIGURE 3
Household statuses based on entering and stay at off–farm and exiting and continuing farming.

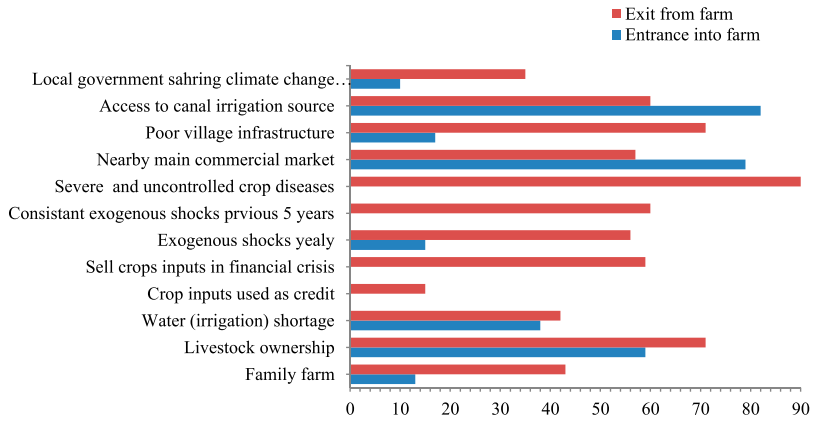


FIGURE 4
Perceived main factors for entrance and exit into and from farming (percentage of households who responded accordingly.

In addition, we tested for multicollinearity among the explanatory variables using the Variance Inflation Factor (VIF). If the maximum VIF value is above 10, there will be econometrically problematic (Wooldridge, 2009; Bai et al., 2010). In our test, the VIF is below 3.5 and therefore there is not found multicollinearity.

5.2 Econometric approach

We used Probit model due to the dichotomous nature of both dependent variables—farm exit and farm entry, whose estimation equations are expressed in Eqs 10, 11, respectively.

$$prob_{(exit)} = \gamma_1 HH_i + \gamma_2 HC_i + \gamma_3 LC_{i+}\gamma_4 NC_i + \gamma_5 EC_i + \gamma_6 ES_i + \gamma_7 CI_i + \gamma_8 LC_i + \gamma_9 \varphi_t + \xi \quad (10)$$

$$prob_{(entry)} = \theta_1 HH_i + \theta_2 HC_i + \theta_3 LC_i + \theta_4 NC_i + \theta_5 EC_i + \theta_6 ES_i + \theta_7 CI_i + \theta_8 LC_i + \theta_9 \varphi_t + \nu \quad (11)$$

Where HH_i is the indicator of household head characteristics such as age, education and status of immigration, HC_i is household characteristics such as fixed inputs of labor and capital and household size. Household wealth is measured by building material of house (concrete or mud) and grain shortage in the last

year. NC_i indicates natural capital (e.g., land ownership, access to canal irrigation, water shortage, water availability for irrigation in Rabi and Kharif seasons[11], actively working Khal Panchayats system[12]and livestock ownership). EC_i indicates economic capital (e.g., owned and run off-farm business, off-farm income, outstanding loans, crop inputs used as credit and for sale in financial crisis). We also estimated total household non-agricultural income (e.g., government transfers, remittances, salaries from off-farm employment, wages, insurance and pensions). ES_i indicates climate shocks (e.g. abnormal temperature, droughts, crop diseases, rainfall and floods). CI_i indicates climate change investment (e.g., distance from MFIs, climate relevant knowledge provided by local government, and access to commercial markets and all-weather road). LC_i indicates locational characteristics (e.g., access to nearby city and to commercial markets, distance to off-farm source (factory/industries) and district, access means to main commercial market, and quality of village infrastructure) (Refer to Supplementary Appendix SA1). γ s and θ s are the corresponding vectors of parameters to be estimated, ξ and ν are the error terms.

5.3 Impact of entry and exit on household wellbeing

We measured wellbeing of a household by three major indicators: total household income, index of food security status, and ability of saving for an emergency fund. We constructed an index of food security status for the year 2012, based on the PRHPS survey that collected information on households' experience of food shortage caused by various climate shocks and illness or death of household member(s) between 2012 and 2014. The index ranges from zero to three, zero being no food shortage and three being the highest level of food shortage. Households' ability of saving for an emergency fund was coded as "1" when they are able to raise Rs.2000[13] during an emergency and "0" otherwise.

11 Pakistan has two major crop seasons: Kharif (broadly July to October) and Rabi (broadly October to March/April).

12 Khal Panchayats or water users' associations are mandated to mediate water distribution conflicts, maintain watercourses, report on tampering of outlets and shortage of water supply in the outlet to minor or distributary–level farmer organizations, collect water charges, and provide timely information about rotational running of channels to the farmers.

13 Rs.2000/day or ($22/day) earning is a reasonable amount for households to survive particularly in case of emergency such as floods, droughts or any disaster occurring in the community.

5.4 Propensity score matching

To evaluate the impact of farm exit and new entrance, we apply the propensity score matching (PSM) method to control for selection bias. An evaluation that failed to control for such selection bias would conflate the effects of farm exit and new entrance on outcomes with the effects of pre-existing differences between farm exit and new entrance. When applying the PSM method, we also test the sensitivity of estimates to potential hidden biases. In theory, the impacts of a farm exit and new entrance should be evaluated by estimating the average treatment effect on the treated (ATT). Thus, we employed PSM to address any self-selection bias of household in their entry/exit decision, because the model matches households that share the same pre-treatment observed socioeconomic characteristics (Heckman et al., 1997; Ali and Peerlings, 2012).

Let $D_i \in \{1, 0\}$ be an indicator whether a household i has received a treatment or not. The propensity score $P(X)$ is defined as the conditional probability of receiving a treatment given pre-treatment characteristics as:

$$P(X) \equiv prob(D_i = |X) = E(D_i|X) \tag{12}$$

where X denotes a vector of pre-treatment characteristics and E is the expectation operator. The propensity score can be predicted with either a logit model under the assumption of a normal or logistic cumulative distribution, respectively. Once the propensity scores are generated, the treatment effect can then be calculated by selecting households that are "nearest neighbor 1-to-1 matching method with replacement" in terms of their estimated propensity scores. The most common estimate of treatment effects in the evaluation literature is the average treatment effect on the treated (ATT). If the potential outcome of the treatment, which is defined as household wellbeing previously, is denoted by $Y_i(D_i)$, then the average treatment effect (ATT) is given as:

$$ATT = E(T|D = 1) = E(Y_1|D = 1) - E(Y_0|D = 1) \tag{13}$$

Where $E(Y_1 D = 1)$ is the expected outcome for those households that have actually received a treatment, in this case those that have entered into or exited from farming, and $E(Y_0|D = 1)$ is the counterfactual for the treated, which estimates what the outcome would be if those households that have in fact received a treatment do not do so. An important assumption of PSM is the Conditional Independence Assumption (CIA), which states that the set of pre-treatment observable characteristics that are included in the matching should determine both the probability of receiving a treatment (entering into and exiting from farming) and the outcome of interest (household wellbeing); that is $(Y_0, Y_1) \perp D|X$, denoting the statistical independence of (Y_0, Y_1), conditional on pre-treatment observable characteristics X (Heckman et al., 1997). Given that the CIA holds, the PSM estimate for the ATT can be written as:

$$ATT_{PSM} = EP_{(X|D=1)}\{E[Y_1|D = 1, P(X)] - [Y_0|D = 1, P(X)]\} \tag{14}$$

TABLE 2 Descriptive analysis and household characteristics comparison (mean and percentage).

Livelihood assets	New entrance (*N* = 226)	Stay at off farm (*N* = 716)	Left farming (*N* = 136)	Continue farming (*N* = 789)
Household head characteristics				
Age (years)	46.42	46.15	44.26**	47.03
Education (years)	3.72	3.56	3.99	3.49
Immigrant (Yes = 1)	0.04	0.04	0.14**	0.07
Household characteristics				
Family size (No.)	6.26	6.18	6.09***	6.84
Number of working age individuals (16–64 years)	2.60	2.38	2.44	2.79
Number of working age children (6–15 years)	1.49*	1.46	1.43**	1.55
Number of elderly persons (>64 years)	0.26	0.20	0.13***	0.28
Pucca (concrete) house (Yes = 1)	0.06*	0.03	0.01**	0.05
HHs faced shortage of grain food during a year (Yes = 1)	0.27*	0.21	0.31***	0.15
Natural capital				
Own land (Yes = 1)	0.13*	0.05	0.43***	0.77
Livestock ownership (Yes = 1)	0.59**	0.50	0.71***	0.87
Canal irrigation (Yes = 1)	0.82***	0.70	0.60***	0.74
Water (irrigation) shortage during the year (Yes = 1)	0.38***	0.52	0.42	0.44
Khal Panchayats system exists (Yes = 1)	0.07***	0.15	—	—
Khal Panchayats system actively work (Yes = 1)	0.15***	0.05	—	—
Water availability in Rabi season (No. of weeks)	—	—	9.49**	11.18
Water availability in Kharif season (No. of weeks)	—	—	10.61***	13.27
Economic capital				
Household owned off-farm business (Yes = 1)	0.15	0.15	0.40	0.37
Household Off-farm income (>Rs. 300 k = 1)	0.27	0.23	0.02	0.01
Household total income (Rs.)	213,792	258,086.3	273,426.5	381,160.6
Outstanding loans (Yes = 1)	0.07**	0.12	0.12*	0.10
Crop inputs used as credit and paid markup (Yes = 1)	—	—	0.15*	0.11
Households sell crops inputs in financial crisis (Yes = 1)	—	—	0.59*	0.32
Climate shocks				
Household affected by natural disasters every year (Yes = 1)	0.15	0.05	0.56***	0.33
Household farms affected by natural disaster shocks for last 5 years (Yes = 1)	—	—	0.60***	0.37
Household farms affected by uncontrolled crop diseases (Yes = 1)	—	—	0.90**	0.68
Climate investment				
Distance to nearest from MFIs (Km)	19.93	18.75	14.27***	17.87
Access to nearest all-weather road (Mins)	49.42	55.22	52.69***	40.85
Local government helping in sharing climate change knowledge (Yes = 1)	0.10	0.27	0.35*	0.02
Locational characteristics				
Nearby city travelling time (minutes)	33.52	35.76	43.74***	33.78
Nearby small commercial market distance (km)	15.05	16.14	20.55**	16.66
Distance to district headquarter (km)	49.20***	43.34	33.76***	42.05
Nearby main commercial market (Pacca_road = 1)	0.79	0.78	0.57	0.61
Poor village infrastructure (kacha road = 1)	0.17	0.17	0.71	0.70
Entry and Exit decisions across provinces				
Punjab	0.14	0.57	0.07	0.50
Sindh	0.06	0.14	0.06	0.23
KPK	0.04	0.05	0.02	0.14
Overall	0.24	0.76	0.15	0.85

Notes: The significance differences between entrance and stay at off-farm, continuing and exiting farming were tested using a one-way ANOVA F-test or a chi-square as appropriate. ***, **, and * are significant at 1%, 5%, and 10% level, respectively.

To eliminate outliers that have very high and very low propensity scores, the matching should be restricted to the area of the common support in the sample, which can be done by dropping the treatment observations at which the propensity score density of the control observation is the lowest (Sianesi, 2004). To be effective, matching should also balance explanatory variables across the treated and non-treated groups. A balancing test performed after the match can check the quality of the match by assessing the extent to which differences in explanatory characteristics between the treated and non-treated groups have been eliminated.

6 Results and analyses

6.1 Descriptive statistics

Table 2 that the age of household heads exiting farming (44) is significantly less than that of those staying farming (47). Household migration status contributes significantly to farm exit (14% vs. 7%), but insignificantly to farm entry. Households that have more family members aged 6 to 15 are significantly more likely to enter farming (1.49 vs. 1.46) and exit farming (1.43 vs. 1.55). This implies that family laborers, particularly working-age youths and adults might prefer to seek off-farm jobs. Similarly, those living in a pucca or concrete house are significantly more likely to enter farming and staying at off-farm (6% vs. 3%), as well to continue and exit farming (5% vs. 1%). The proportion of households who faced a grain food shortage during a year has a significant influence on the decision to enter into farming (27% vs. 21%) and exit from farming (31% vs. 15%). Having a family farm or land ownership, is significantly more important for entry decisions, and less important when it comes to exit decisions (13% vs. 43%), respectively. This result is surprising for exiting households with land ownership (family farm), and raises a question about future of family farm in the country. However, in case of farm exit despite land ownership the reasons may be unavailability of successors within family or such farmers faced consistent crop losses due to climate shocks and severe crop diseases, lower output prices, and higher input prices, and therefore they might decide to rent out their land and shift to off-farm activates. Furthermore, the results depict that livestock ownership is also contributing significantly for both entry and exit decisions (59% vs. 71%), respectively. Note that new entrant and households working off-farm may still own livestock by sharing with friends/relatives or landlords and animal feed is provided by owners of fodder, which is a common practice in rural Pakistan. Alternatively, women may take responsibility for the rearing of animals in rural Pakistan, and undertake field work such as crop sowing, harvesting, and bring herbs and grass as fodder for the animals (Ahmad and Ma, 2020b). Access to irrigation canal and Khal Panchayat system plays a significant role in entry decisions, whereas water shortage during crop seasons play a significant role in exit decisions. Economic capital in terms of outstanding loan, crop inputs used as credit and sold by farmers on net cash in financial crisis significantly increase farm exit (Ahmad and Ma, 2020b).

Turning to climate shocks, crops affected by natural disasters and severe crop diseases play a significant role in exit decisions. Furthermore, climate investments - the distance to MFIs and all-weather roads—play a significant role in exit farming.

For locational characteristics, distances from off-farm source location and home district are significantly contributing to both entry and exit decisions. Additionally, travelling time to nearby city and distance to commercial markets also significantly contribute to farm entry and exit.

6.2 Household head characteristics

The Probit models were then used to estimate the probability of farm entry and farm exit, respectively. Table 3 presents the marginal effects from the probit regression for decisions on farm entry and exit. In the exit model, migration status plays a significant role. Age makes a significant difference when it comes to farm entry, particularly in developed countries where farmers' retirement plans involve the recruitment of new and younger farmers (Kimhi and Bollman, 1999; Pietola et al., 2002; Vare and Heshmati, 2004; Glauben et al., 2006). Our results suggest that as the age of the household head increases, households are more likely to enter into farming. But age is insignificant in the exit model. This result seems plausible as the elderly may move to the farming sector when they become less productive with off-farm work due to aging and declined health conditions.

6.3 Household characteristics

In the entry model, larger households are less likely to enter into farming, mainly because such households have "surplus" labor to generate sufficient off-farm income. It also found that the households having more members in working age (16–64) are more likely to enter into farming. This result suggests that the presence of more adults overcome labor constraints and provide more hands in fertilizing the crops, weeding, taking out infested plants, and transplanting and harvesting. This result partly supports Ahmad et al. (2020) who found that the significant and positive relationship between working-age family member and farm exit. Surprisingly, households experiencing food shortage were more likely to exit from farming, which raises a serious concern in relation to why they faced food shortage despite growing their own grain. This seems to be contradictory to the approach of the Millennium Development Goal (MDG) that prioritizes farm activities as a means to reduce poverty and hunger (World Bank, 2008). Possibly this is because these households had to sell all their grain output to repay the previous loans or experienced severe crop losses due to climate shocks (Ahmad and Ma, 2020a).

TABLE 3 Marginal effects of probit regression for the probability of entrance and exit into and from farming.

Livelihood assets	Probability of entrance	Probability of exit
Household head characteristics		
Age (years)	0.002* (0.001)	0.001 (0.001)
Education (years)	0.002 (0.003)	0.002 (0.002)
Immigrant (Yes = 1)	−0.101 (0.068)	0.075* (0.039)
Household characteristics		
Family size (No.)	−0.019* (0.009)	-0.012 (0.008)
Number of working age individuals (16–64 years)	0.033** (0.013)	0.006 (0.011)
Number of working age children (6–15 years)	0.020 (0.014)	0.002 (0.012)
Number of elderly persons (>64 years)	0.045 (0.028)	−0.032 (0.024)
Pucca (concrete) house (Yes = 1)	0.122* (0.067)	−0.134* (0.074)
Household faced shortage of grain food during a year (Yes = 1)	−0.008 (0.059)	0.055** (0.025)
Natural capital		
Own land (Yes = 1)	0.409*** (0.081)	−0.136*** (0.022)
Livestock ownership (Yes = 1)	0.049* (0.026)	−0.074*** (0.025)
Canal water irrigation (Yes = 1)	0.272*** (0.042)	−0.190*** (0.017)
Water shortage during seasons (Yes = 1)	-0.146*** (0.030)	0.146*** (0.043)
Khal Panchayats system exists (Yes = 1)	-0.135*** (0.046)	—
Khal Panchayats system work actively (Yes = 1)	0.134*** (0.050)	—
Water availability in Rabi season (No.of weeks)	—	0.004* (0.002)
Water availability in Kharif season (No.of weeks)	—	−0.006** (0.002)
Economic capital		
Household owned off-farm business (Yes = 1)	−.015 (0.036)	−0.027 (0.043)
Household Off-farm income (>Rs. 300 k = 1)	0.001 (0.034)	0.172** (0.078)
Outstanding loans (Yes = 1)	−0.094* (0.044)	0.017 (0.032)
Households sell crops inputs in financial crisis (Yes = 1)	—	0.353*** (0.023)
Crop inputs used as credit (Yes = 1)	—	0.113*** (0.028)
Climate shocks		
Household affected by natural disasters every year (Yes = 1)	−0.138*** (0.044)	0.092* (0.050)
Household farms affected by natural disaster shocks for last 5 years (Yes = 1)	—	0.094* (0.049)
Household farms affected by uncontrolled crop diseases (Yes = 1)	—	0.148*** (0.034)
Climate investment		
Distance to nearest from MFIs (Km)	0.0003 (0.002)	−0.007*** (0.002)
Access to nearest all-weather road (Minutes)	−0.0003 (0.001)	−0.002** (0.001)
Local government sharing well climate change knowledge (Yes = 1)	—	−0.085* (0.092)
Locational characteristics		
Nearby city travelling time (minutes)	−0.001 (0.001)	0.003*** (0.001)
Nearby small commercial market distance (Km)	0.002 (0.002)	0.004*** (0.001)
Off-farm source (factory/ industries) distance from village less than 20 km = 1)	0.002 (0.002)	0.061** (0.028)
Distance to district headquarter (km)	−0.001 (0.001)	−0.001 (0.001)
Nearby main commercial market distance (Pacca_road = 1)	0.037 (0.037)	−0.104*** (0.027)
Poor village infrastructure (kacha road = 1)	—	−0.093*** (0.026)
Number of observations	942	925
Chi squared	213***	264.49***
Pseudo R^2	0.205	0.3424
Log likelihood	−412.55	−253.95

Notes: Robust standard errors are reported in parenthesis. ***, **, and * are significant at 1%, 5%, and 10% level, respectively.

6.4 Natural capital

Access to capital as land ownership (family farm) increases the probability of farm entry by 40%, and reduces the probability of farm exit by 13%. Land ownership (family farms) are the most common business model in small-scale agriculture (Davidova and Thomson, 2013). Consequently, land ownership plays an important role in farming decisions. For example, in the case of access to credit or agricultural loans, only landowners can benefit from these services by using land as collateral. Households who owned land (family farm) and access to canal source irrigation were less likely to exit from farming with probabilities of 13% and 19%, respectively. As it may be the reason that family farms and water availability make farmers stay farming without extra cost of purchasing or hiring land and pumping groundwater for irrigation.

Livestock ownership encourages households to stay farming or enter into farming (as mixed-crop livestock production system). Ownership of livestock increases and reduces the probability of farm entry and farm exit by 5% and 7%, respectively. Mixed-crop livestock production is an integral part of farming all over the world including Pakistan, and is closely linked to livelihood strategies as a major source of food and income (Gurung, 1987; Herrero et al., 2010; Ahmad and Ma, 2020b). Households that own livestock (such as buffalo, cattle, sheep, and goats etc.) were significantly more likely to enter into farming and significantly less likely to exit farming. In fact, approximately 35%–40% of the

Pakistani population are dependent on livestock as their main source of income and the livestock sector provides food for over 8 million rural families (Pakistan Bureau of Statistics, 2019).

Households who have less access to water in Rabi crop season were more likely to exit farming compared with Kharif season. This is because most of the farms have enough canal irrigation in Kharif season while have limited water availability in Rabi season. Hence households exited farming due to irrigation water shortage in Kharif season. In contrast, households that have better access to canal source[14] of irrigation were more likely to enter farming, with an increase in the probability by 27%, less likely to exit farming, with a decrease in the probability by 19%. Farmers seem to prioritize the use of canal water, as it is the cheapest source of irrigation and it could also encourage farmers to enter farming. Furthermore, we found that where the Khal Panchayats system was working actively, farmers were more likely to enter into farming, increasing the probability by 13%. Due to shortage of water during the crop seasons, it reduces and increases the probability of entrance and exit decision by 14% and 15%, respectively. Additionally, our results in the exit model demonstrate the importance of irrigation across seasons.

6.5 Economic capital

As annual income generated from off-farm activities increases (more than Rs. 300 k), it also increases the probability of farm exit by 17%. Thus, household who worked at both on-farm and off-farm can compare both income and decide either to stay or exit farming. Hence, those who left farming, might find farming as a less rewarded occupation and eventually they could decide to exit farming due to higher off-farm income. However, in the entrance model, influence of off-farm income is positive but insignificant.

Financial constraints, especially those related to the use of crop inputs (rising prices of fertilizers, seed, pesticides, diesel for pumping ground water, etc.), seem to have serious effects on farmers and encourage them to exit farming in Pakistan. In this case, farmers experience large mark-ups on these input costs when they use them on credit, making them have no option but utilise their own income and limited (or non-existent) savings. As a result, outstanding loans create additional pressure for farmers, which affect their farming and even sometimes lead farmers to committing suicide (Mishra, 2006; Gruère and Sengupta, 2011). Our results show that farming households that use crop input as credits are more likely to exit farming, which increases the probability by 11%. Indeed, our results show that outstanding loans or debts can significantly increase the probability of farm exit. Meanwhile, longstanding loan and debt pressure could distress farmers and make them dislike or exit farming (Deshpande and Prabhu, 2005). Thus, as expected any mark-up imposed on using crop inputs (fertilizer, pesticide, seeds, etc.) as credit could push farmers to exit from farming due to financial constraints.

Our results also suggest that farm households who sold their crop inputs in financial crisis to feed families, were more likely to exit from farming, by a probability of 35%, which is higher than the effect of any other factors in this study. Approximate 60% of households have exited farming activities between 2012 and 2014, with one of the most important reasons being that farmers sold their borrowed or credit crops inputs (mainly fertilizer) in financial crisis (Ahmad and Ma, 2020b). For example, farmers first borrowed or credited crop inputs from input dealers at huge mark-ups, and then, they sold those crop inputs to neighbour farmers (someone else) or other input dealers on net cash at lower than market prices to deal with an emergency and feed families in financial crises. As a result, this behaviour not only increased the burden of loans on farmers but also converted their previous input-driven small loans into larger loans, if they fail to pay when the harvest was completed. To conclude, farmers should not depend only on on-farm income, and they should be involved with part time off-farm work to stabilize income and support their families in case of financial crisis particularly during the crop growing stages. As the crop input mark-ups are higher, first, farmers should avoid using these crop inputs on credit, and second should also avoid selling heavy mark-up crop inputs during an emergency, because this will create extra burden and lead to farm exit in the end. In addition, this behaviour could also discourage farm entry.

6.6 Climate shocks

Erratic climate is severely affecting the livelihoods of households who depend upon agricultural production. The results show that climate weather shocks not only affect farm exit but also have negative impact on farm entry. In fact, households who live in disasters prone regions were less likely to enter farming, with the probability being lowered by 14%. Similarly, households who faced climate shocks during last consecutive five years and were affected by severe crop disease were more likely to exit farming, with the probabilities of farm exit being increased by 9% and 15%, respectively.

To conclude, households that have experienced large crop losses due to heavy rain-fall, floods, droughts and severe crop disease are more likely to exit from farming, and these climate shocks push households to diversify their livelihoods beyond agriculture. As a result, the productivity of the agricultural sector decreased gradually due to emerging high input prices, lower output prices, water shortages, and climate shocks. However, those living in these regions face a number of challenges, such as food insecurity and poverty, driven in part by climate shocks, which encourage them to seek alternatives to farming and diversify their livelihoods towards off-farm activities (Glauben et al., 2006; Bhandari, 2013). Thus, climate shocks make farming livelihood more vulnerable and increases the likelihood to diversification. This finding is consistent with those of other studies in Pakistan on climate

14 In Pakistan, the availability of canal water supply is inconsistent (only 4–6 months in a year).

change and climate shocks and their effects on survival of livelihoods (Ahmad et al., 2020; Ahmad and Ma, 2020b).

6.7 Climate change investment

We found significant effects for distance to nearest MFIs, access to nearest all-weather road and the role of local government in farmers' wellbeing (only in exit model). The availability of MFIs in nearby locations has no significant effect on the entry decision. This may be because the poor access to MFIs reduces the required capital to start farming, and enables households to use their capital endowment for investment without necessarily being credit constrained (Huang et al., 2008; Ruan and Zhang, 2009; Ali et al., 2010). For example, those who have access to agricultural loans are more likely to continue farming while access to agricultural loans acts as an incentive for new entrants into farming. It appears that rural finance is important to farmers and therefore policymakers should consider policies to resolve such financial constraints to attract new farmers. In fact, the role of local government, particularly knowledge sharing of climate change and livelihood diversification, could significantly reduce the probability of farm exit by 8%.

6.8 Locational characteristics

The results show that the nearby city travelling time, distance to nearby small commercial markets, distance to nearby off-farm source (factory/industries), distance to main commercial markets, poor village infrastructure are all found significant in exiting from farming except for the distance to district headquarter. These results suggest that rural finance and urban employment could create significant and positive associations on entry into farming for all those who want to continue farming, or who are thinking of becoming farmers. Households that live within additional community service areas have a better chance of engaging in off-farm opportunities, which could induce them to start off-farm work. The distance to nearby off-farm sources (factories, mills, and industry) has a positive and significant association on farm exit. On the other hand, the distance to a nearby small commercial market has a positive and significant association on farm exit, but an insignificant association on farm entry.

6.9 Farm entry-exit decisions and household wellbeing

There are different matching methods to calculate the average treatment effects in the evaluation literature. The one we used in this study is before and after the nearest neighbor 1-to-1 matching with replacement, which associates the outcome of the treated household with the matched outcome that is given by this 1-to-1 matching method and weighted average of all the non-treated households. Because the weighted average of all the non-treated households is used to construct the counterfactual outcome, 1-to-1 matching method has an advantage of lower variance (Heckman et al., 1998).

A *t*-test was used to compare the mean of each covariate between the treatment and control group after the matching procedure. If the matching was successfully accomplished, the mean difference after matching should be insignificant. The results of the *t*-test showed that the differences in the covariates became insignificant after the matching procedure, which indicates that the observable characteristics of the control group were sufficiently similar to those of the treatment group after matching. The matching quality tests for the entry and exit models suggest that the matching procedures have performed well in terms of avoiding systematic difference in the distribution of pre-treatment observable covariates that are included in the PSM between the treated and non-treated groups.

To check the above results of the match are robust, a sensitivity analysis is performed by using a nearest neighbor 1-to-1 matching method with replacement. The findings confirm that the matching results are quite robust. Although the above results of the PSM indicate that biases from observables are controlled, it might be difficult to infer a causal relationship between diversification and wellbeing as there could still be some unobserved factors that exert certain effects on both farm entry (exit) and households' wellbeing. In addition, farm and non-farm earnings can reinforce each other, which could then influence households' wellbeing through indirect channels such as tightening of the agricultural labor market or raising demand for agricultural products, etc. (Janvry de, 1994; Loening and Mikael, 2009).

7 Propensity scores, balance tests and sensitivity analysis

The characteristics of exit and continue farming households are shown in Table 4. We find that the difference between exit and continue farming was statistically significant in household owned off-farm business, owned land (family farm), access to canal irrigation source and poor village infrastructure. Regarding the household owned off-farm business, exit farmers significantly were less (5%) compared to continued farms (8%). Similarly, we found that the household that exit farms were also those who had family farm (owned land), anyhow these were less (43%) compared to continued farms (77%). We also observe that canal source of irrigation

TABLE 4 Variables, definitions, means, and difference-in-means tests.

Variables	Mean all (N = 925)	Left farming (N = 136)	Continue farming (N = 789)
Education (years)	3.563 (4.430)	3.985 (5.121)	3.490 (4.299)
Number of working age individuals (16–64 years)	2.741 (2.049)	2.441 (1.877)	2.792 (2.073)
Immigrant (Yes = 1)	0.082 (0.275)	0.140 (0.348)	0.072 (0.259)
Household owned off-farm business (Yes = 1)	0.078 (0.268)	0.051 (0.222)**	0.082 (0.275)
Own land (Yes = 1)	0.722 (0.448)	0.426 (0.496)***	0.773 (0.419)
Canal water irrigation (Yes = 1)	0.719 (0.450)	0.596 (0.493)***	0.740 (0.439)
Household affected by natural disasters every year (Yes = 1)	0.366 (0.482)	0.559 (0.498)***	0.333 (0.472)
Crop inputs used as credit (Yes = 1)	0.332 (0.471)	0.331 (0.472)	0.332 (0.471)
Water shortage during seasons (Yes = 1)	0.436 (0.496)	0.419 (0.495)	0.439 (0.497)
Poor village infrastructure (kacha road = 1)	0.256 (0.437)	0.140 (0.348)***	0.276 (0.447)

Notes: Numbers are means; numbers in parentheses are S.D., values. ***, **, and * are significant at 1%, 5%, and 10% level, respectively.

TABLE 5 Variables, definitions, means, and difference-in-means tests.

Variables	Mean all (N = 942)	New entrance (N = 226)	Stay at off farm (N = 716)
Education (years)	3.597 (4.328)	3.717 (4.254)	3.559 (4.353)
Number of working age individuals (16–64 years)	2.431 (1.728)	2.602 (1.705)**	2.377 (1.733)
Immigrant (Yes = 1)	0.039 (0.194)	0.040 (0.196)	0.039 (0.194)
Household owned off-farm business (Yes = 1)	0.149 (0.356)	0.146 (0.354)	0.149 (0.357)
Own land (Yes = 1)	0.034 (0.181)	0.128 (0.335)***	0.004 (0.065)
Canal water irrigation (Yes = 1)	0.728 (0.445)	0.823*** (0.383)	0.698 (0.459)
Household affected by natural disasters every year (Yes = 1)	0.076 (0.266)	0.155 (0.363)***	0.052 (0.222)
Water shortage during seasons (Yes = 1)	0.486 (0.500)	0.376 (0.485)***	0.521 (0.500)
Outstanding loans (Yes = 1)	0.109 (0.312)	0.071 (0.257)*	0.122 (0.327)
Poor village infrastructure (kacha road = 1)	0.167 (0.373)	0.168 (0.375)	0.166 (0.373)
Khal Panchayats system work actively (Yes = 1)	0.076 (0.266)	0.150 (0.358)*	0.053 (0.224)
Local government sharing well climate change knowledge (Yes = 1)	0.717 (0.451)	0.721 (0.449)	0.715 (0.452)
Nearby main commercial market distance (Pacca_road = 1)	0.782 (0.413)	0.788 (0.410)	0.781 (0.414)

Notes: Numbers are means; numbers in parentheses are S.D., values. ***, **, and * are significant at 1%, 5%, and 10% level, respectively.

significantly associated with farm exit (60%), which is lower than farm continuing households (74%). More important, natural disasters significantly increases the percentage of exit farm (56%) which was higher than continued farming (33%). Furthermore, village infrastructure also significantly associated with farm exit, the results show that 14% households exit farming due to poor village infrastructure particularly connection of roads which is mud or kacha road.

Table 5 represents the characteristics of new entrance into farming. We find that the difference between new entrance and stay at farm was statistically significant and higher in new entrance (2.602) compared to stay at farm (2.377). Similarly, we find that the new entrance farms were also those who had owned family farm (owned land) were higher (16%) compared to stay at farm farms (5%). We also observe that canal source of irrigation significantly associated with new entrance and attract new entrance as 82% household enterd into farming due to canal irrigation source. More important, natural disasters significantly associated with new entrance by 15%. Furthermore, Khal Panchayats system significantly attracts 15% new entrance. More important, household affected by natural disasters significantly increases the percentage of exit farm by 56% which was higher than stay at farm farming (33%).

Table 6 presents the logit regression to generate the propensity scores. The goodness of fit can be measured by the pseudo R^2 value, and logit estimation gives a pseudo R^2 of 0.158 in farm exit model. The results indicate that, all other

TABLE 6 Logit regression results (dependent variables are whether the farm exit farm).

Variables	Coef	S.E
Education (years)	0.03	0.02
Number of working age individuals (16–64 years)	−0.03	0.06
Immigrant (Yes = 1)	0.56*	0.32
Household owned off-farm business (Yes = 1)	−0.34	0.44
Own land (Yes = 1)	−1.51***	0.22
Canal water irrigation (Yes = 1)	−1.28***	0.28
Household affected by natural disasters every year (Yes = 1)	0.87***	0.21
Crop inputs used as credit (Yes = 1)	−0.37**	0.22
Water shortage during seasons (Yes = 1)	0.56*	0.27
Poor village infrastructure (kacha road = 1)	−1.04***	0.29
Constant	−0.338***	0.293
LR. chi2	122.13	
P > chi2	0.000	
Log likelihood	−325.136	
Pseudo R^2	0.158	

Note: ***, ** and * are significant at 1%, 5%, and 10% level, respectively.

TABLE 7 Logit regression results (dependent variables are whether the new entrance farm).

Variables	Coef	S.E
Education (years)	0.016	0.020
Number of working age individuals (16–64 years)	0.077	0.047
Immigrant (Yes = 1)	−0.262	0.443
Household owned off-farm business (Yes = 1)	−0.125	0.243
Own land (Yes = 1)	3.086***	0.635
Canal water irrigation (Yes = 1)	1.226***	0.233
Household affected by natural disasters every year (Yes = 1)	0.826***	0.285
Water shortage during seasons (Yes = 1)	−1.265***	0.200
Outstanding loans (Yes = 1)	−0.807*	0.314
Poor village infrastructure (kacha road = 1)	0.176	0.233
Khal Panchayats system work actively (Yes = 1)	0.922***	0.301
Local government sharing well climate change knowledge (Yes = 1)	0.050	0.186
Nearby main commercial market distance (Pacca_road = 1)	0.171	0.218
Constant	−2.117***	0.332
LR. chi2	146.30	—
P > chi2	0.0099	—
Log likelihood	−445.874	—
Pseudo R^2	0.141	—

Note: ***, ** and * are significant at 1%, 5%, and 10% level, respectively.

things being equal in our sample, migration, land ownership, irrigation source (canal), natural disasters, crop inputs used as credit, water shortage during seasons and poor village infrastructure are more likely to be exit farming. Similarly, Table 7 also presents the logit regression to generate the propensity scores and logit estimation gives a pseudo R^2 of 0.141 in new entrance model. The results indicate that, all other things being equal in our sample, land ownership, irrigation source (canal), natural disasters, water shortage during seasons and Khal Panchayats system work actively are more likely to be entrance into farming.

Tables 8, 9 show the results of balancing tests for the PSM with before and after nearest neighbor 1-to-1 matching method (refer to Supplementary Appendix SA2, Supplementary Appendix SA3, Supplementary Appendix SA4). Further we used *t*-test to compare the mean of each covariate between the treatment and control group after the matching procedure. If the matching was successfully accomplished, the mean difference after matching should be insignificant. The results of the *t*-test showed that the differences in most of the covariates became significant after the matching procedure, which indicates that the observable characteristics of the control group were sufficiently similar to those of the treatment group after matching.

Table 10 presents household wellbeing results for entry and exit for three outcome variables (Total household income, Domestic food shortage, and Able to earn money in an emergency) by using nearest neighbor 1-to-1 matching method. The results provide strong evidence that entry decision has significantly and negatively associated with household wellbeing in terms of total household income, whereas positively associated with Domestic food shortage and Able to earn money in an emergency but results are insignificant. The results indicated that households that have entered into farming on average have an annual income Rs. 39,340.8 ($409.79) less than those who have not entered into farming. Furthermore, we find that the standard matching ATTs of Domestic food shortage is positive and statistically insignificant, indicating that entering into farming results in increase in food shortage. However, these results void our hypothesis that entry into farming could be an excellent effort for hushed to be self-sufficient in food security in the future. Turning to farm exit, the ATTs of Able to earn money in an emergency is statistically significant and indicates that households that have exit farming on average have an annual income Rs. 11,716.63 ($122.04) more than those who have not exit farming but the results are insignificant. Furthermore, we find that the standard matching ATT of Domestic food shortage is positive and statistically significant, indicating that exit from farming results increase in food shortage. However, these results are in line towards our hypothesis that exit farming could lead household to be food insecure in the future.

TABLE 8 Results of balancing tests before and after the nearest neighbor 1-to-1 matching with replacement (For exit farming).

Variables	Sample	Mean				*t*-test		V(T)/(VC)
		Treated	Control	% Bias	% Redu. Bias	t	$p>\|t\|$	
Education (years)	Unmatched	3.99	3.49	10.5	—	1.2	0.229	1.42*
	Matched	4.01	3.42	12.7	−21.1	1.02	0.310	1.27
Number of working age individuals (16–64 years)	Unmatched	2.44	2.79	−17.7	—	−1.85	0.065	0.82
	Matched	2.45	2.34	5.8	67.1	0.54	0.592	1.30
Immigrant (Yes = 1)	Unmatched	0.14	0.07	22	—	2.65	0.008	1.80*
	Matched	0.14	0.19	−16	27.3	−1.08	0.280	0.79
Household owned off-farm business (Yes = 1)	Unmatched	0.05	0.08	−12.4	–	−1.24	0.214	0.65*
	Matched	0.05	0.04	6.5	47.3	0.65	0.514	1.43*
Own land (Yes = 1)	Unmatched	0.43	0.77	−75.5	—	−8.66	0.000	1.40
	Matched	0.43	0.39	9.2	87.7	0.71	0.480	1.03
Canal water irrigation (Yes = 1)	Unmatched	0.60	0.74	−31	—	−3.48	0.001	1.26
	Matched	0.60	0.67	−15.8	49.1	−1.26	0.210	1.09
Household affected by natural disasters every year (Yes = 1)	Unmatched	0.56	0.33	46.5	—	5.11	0.000	1.12
	Matched	0.56	0.54	4.2	90.9	0.34	0.736	0.99
Crop inputs used as credit (Yes = 1)	Unmatched	0.33	0.33	−0.3	—	−0.03	0.978	1.00
	Matched	0.33	0.36	−6.1	−2,320	−0.49	0.623	0.96
Water shortage during seasons (Yes = 1)	Unmatched	0.42	0.44	−3.9	—	−0.42	0.674	0.99
	Matched	0.41	0.45	−7.9	−102.2	−0.65	0.517	0.98
Poor village infrastructure (kacha road = 1)	Unmatched	0.14	0.28	−34.1	—	−3.39	0.001	0.60*
	Matched	0.14	0.13	1.7	95.1	0.16	0.874	1.04

Note: ***, **, and * are significant at 1%, 5%, and 10% level, respectively.

Furthermore, we calculated the critical value of $\Gamma^{\dagger}$ (Table 10). For the significant ATTs, the value of $\Gamma^{\dagger}$ total household income is 1.2. It implies that matched entry farmers with the same observed covariates would have to differ in terms of unobserved covariates by a factor of 1.2 for total household income inference of a significant treatment effect. Similarly, in case of farm exit decision, the value of $\Gamma^{\dagger}$ for Able to earn money in an emergency is 1.6. It implies that matched exit farmers with the same observed covariates would have to differ in terms of unobserved covariates by a factor of 1.6 for able to earn money in an emergrncy to invalidate the inference of a significant treatment effect. Therefore, we conclude that the impact estimates are fairly robust to potentially hidden bias. Apel et al. (2010) reported that the estimation results in applied research often become sensitive to Γ value as small as 1.15. However, the results are sensitive to unobserved characteristics of other insignificant ATTs.

8 Conclusion and implications

This article has investigated and identified factors that affect households' farm entry and farm exit based a dataset of 1867 households. This study also investigates the impact of entry and exit decisions on households' wellbeing by using PSM approach for three wellbeing outcomes. This study has brought fresh insights into sustaining rural livelihoods of both the farming and off-farm sectors.

Firstly, our results suggest that household decision of entry into farming significantly decreases household income, while household decision of exit from farming significantly increases domestic food shortage. This finding indicates that farming would be rural household income source in Pakistan. This finding might explain why a larger share (24%) of rural off-farm working labor has shifted into farming as new entrants in Pakistan since 2014.

Secondly, we found that climate shocks could significantly affect farm entry and farm exit decisions, respectively. It is meant that climate shocks could change farmers' future attitudes towards farming. This finding points out a potentially concern as Pakistan agricultural production and livelihoods are particularly vulnerable to climate shocks. In fact, as stated previously, Pakistan's agriculture sector has been faced serious challenges of water shortages and natural disastrous since 2001, and even currently massive floods are still in this country.

TABLE 9 Results of balancing tests before and after the nearest neighbor 1–to–1 matching with replacement (For new entrance into farming).

Variable	Sample	Mean				*t*–test		V(T)/(VC)
		Treated	Control	% Bias	% Redu. Bias	t	$p>\|t\|$	
Education (years)	Unmatched	3.717	3.559	3.7	—	0.48	0.632	0.95
	Matched	3.689	3.726	−0.9	76.7	−0.09	0.929	0.99
Number of working age individuals (16–64 years)	Unmatched	2.602	2.377	13.1	—	1.71	0.088	0.97
	Matched	2.590	2.617	−1.6	87.9	−0.15	0.88	0.74*
Immigrant (Yes = 1)	Unmatched	0.040	0.039	0.4	—	0.05	0.961	1.02
	Matched	0.042	0.018	12.3	−3255.8	1.44	0.15	2.25*
Household owned off–farm business	Unmatched	0.146	0.149	−1	—	−0.13	0.9	0.98
	Matched	0.142	0.123	5.1	−430.4	0.55	0.582	1.12
Own land (Yes = 1)	Unmatched	0.128	0.004	51.4	–	9.38	0	26.89*
	Matched	0.071	0.059	4.9	90.5	0.49	0.623	15.48
Canal water irrigation (Yes = 1)	Unmatched	0.823	0.698	29.5	—	3.7	0	0.69*
	Matched	0.811	0.797	3.3	88.7	0.36	0.716	0.95
Household affected by natural disasters every year (Yes = 1)	Unmatched	0.155	0.052	34.3	—	5.16	0	2.68*
	Matched	0.137	0.166	−9.8	71.6	−0.84	0.401	0.85
Water shortage during seasons (Yes = 1)	Unmatched	0.376	0.521	−29.4	–	−3.82	0	0.94
	Matched	0.392	0.383	1.7	94.2	0.18	0.859	1.01
Outstanding loans	Unmatched	0.071	0.122	−17.2	–	−2.13	0.033	0.62*
	Matched	0.075	0.065	3.6	79.1	0.43	0.67	1.15
Poor village infrastructure (kacha road = 1)	Unmatched	0.168	0.166	0.5	—	0.07	0.946	1.01
	Matched	0.175	0.156	4.9	−848	0.51	0.611	1.09
Khal Panchayats system work actively (Yes = 1)	Unmatched	0.150	0.053	32.6	—	4.86	0	2.55*
	Matched	0.123	0.097	8.4	74.1	0.83	0.408	1.22
Local government sharing well climate change knowledge (Yes = 1)	Unmatched	0.721	0.715	1.4	—	0.18	0.858	0.99
	Matched	0.717	0.724	−1.7	−21.3	−0.17	0.864	1.02
Nearby main commercial market distance (Pacca_road = 1)	Unmatched	0.774	0.750	1.7	—	0.22	0.827	0.98
	Matched	3.717	3.559	5.8	−247.2	0.58	0.565	0.93

Note: ***, **, and * are significant at 1%, 5%, and 10% level, respectively.

TABLE 10 Average treatment effect on ATTs and critical value of Rosenbaum's Γ by using 1-to-1 matching method.

PSM	Variables	Mean treated	Mean control	ATT	S.E.	$\Gamma^{\dagger}$
Farm Entry	Total household income	213,792	258,086	−39340**	21,598	1.2
	Domestic food shortage	0.231	0.248	0.074	0.047	1.0
	Able to earn money in emergency	0.044	0.048	0.003	0.021	1.3
Farm Exit	Total household income	274,526	26,280,99	11,716	51,038	1.1
	Domestic food shortage	0.311	0.205	0.105*	0.047	1.6
	Able to earn money in emergency	0.037	0.079	−0.042	0.032	1.4

Note: ***, **, and * are significant at 1%, 5%, and 10% level, respectively.

Thirdly, descriptive statistics show that approximate 60% of smallholders exited from farming between 2012 and 2014, and 15% of smallholders exit farming after 2014, because they had to sell off their borrowed or credit crop inputs (mainly fertilizer). The major reason is due to the lack of national macro agricultural support policies. In fact, there are more than 80% are samllholders, but the agricultural subsidies (e.g., fertilizer and machinery purchase) are only provided

for large farms (over 12 acres of land) in Pakistan (Ali et al., 2019).

Gernally, agriculture is still major income source for most of rural huseholds and that is why smallholders exist in Pakistan on the one hand. On the other hand, off-farm income also reduces smallholders to exit farming. More importantly, climate shocks and national agricultural support policy can be the crucial factors for smallholders whether to enage in agricultural production.

As with most research, this study has some limitations. For example, we did not discuss whether any household member inherited family farms as a successor or not. We are also unable to identify whether the entrants into farming were permanent or temporary. These questions are also potential for future research.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

MA: Data processing, making draft, revision; LO: Editing, comments; HM: Supervising, comments, editing, and revision; RL: Comments, revision.

Funding

National Natural Science Foundation of China (Grant No. 71403082); National Social Science Foundation of China (Grant No. 14BGL093).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fenvs.2022.857082/full#supplementary-material

References

Abbasi, S. S., and Kim, T. K. (2018). The factors of migration in Northersn mountain region of Pakistan (A case study of Tehsil Murree, district Rawalpindi. Jordan. *J. Econ. Sci.* 5 (1), 67–75.

Ackoff, S., Bahrenburg, A., and Shute, L. L. (2017). *Building a Future with Farmers: Results and recommendations from the national Young Farmer survey*. Hudson, NY: National Young Farmers Coalition.

ADAS (2004). *Entry to and exit from farming in the United Kingdom: Final report*. London: DEFRA, University of Plymouth, Queen's University Belfast.

Aftab, M., Iftikhar, M., Khan, G. A., and Siddiqui, M. T. (2021). Youth perception and participation in agriculture in district Faisalabad, Punjab, Pakistan. *Int. J. Agric. Ext.* 9, 493–499. doi:10.33687/ijae.009.03.3843

Agarwal, B., and Agrawal, A. (2017). Do farmers really like farming? Indian farmers in transition. *Oxf. Dev. Stud.* 45 (4), 460–478.

Agarwal, B. (2014). *Food security, productivity and gender inequality. Handbook of food, politics and society*. New York, NY: Oxford University Press.

Ahearn, M. C., and Newton, D. (2009). *Beginning Farmers and Ranchers*. Washington, D.C., United States: USDA Economic Research Service. Bulletin Number 53.

Ahearn, M. C. (2011). Potential challenges for beginning farmers and Ranchers. *Choices*. Quarter 2.

Ahmad, M. I., and Ma, H. Y. (2020b). An investigation of the targeting and allocation of post-flood disaster aid for rehabilitation in Punjab, Pakistan. *Int. J. Disaster Risk Reduct.* 44, 101402. doi:10.1016/j.ijdrr.2019.101402

Ahmad, M. I., and Ma, H. Y. (2020a). Climate change and livelihood vulnerability in mixed crop–livestock areas: The case of province Punjab, Pakistan. *Sustainability* 12, 586. doi:10.3390/su12020586

Ahmad, M. I., Oxley, L., and Ma, H. Y. (2020). What makes farmers exit farming: A case study of Sindh province, Pakistan. *Sustainability* 12, 3160. doi:10.3390/su12083160

Ali, A., Rahut, D. B., and Imtiaz, M. (2019). Affordability linked with Subsidy: Impact of fertilizers Subsidy on household Welfare in Pakistan. *Sustainability* 11, 5161. doi:10.3390/su11195161

Ali, H., Shafi, M. M., and Siraj, M. (2014). Determinants of off-farm employment among small farm Holders in rural areas of district Mardan. *Sarhad J. Agric.* 30 (1), 145–150.

Ali, M., and Peerlings, J. (2012). Farm households and nonfarm activities in Ethiopia: Does clustering influence entry and exit? *Agric. Econ.* 43 (3), 253–266. doi:10.1111/j.1574-0862.2012.00580.x

Ali, M., Peerlings, J., and Zhang, X. (2010). *Clustering as an organizational response to capital market inefficiency: Evidence from handloom enterprises in Ethiopia*. Washington, D.C., USA: International Food Policy Research Institute. Discussion Paper Series: 01045.

Apel, R., Blokland, A. A. J., Nieuwbeerta, P., and van Schellen, M. (2010). The impact of imprisonment on marriage and divorce: A risk set matching approach. *J. Quant. Criminol.* 2, 269–300. doi:10.1007/s10940-009-9087-5

Bai, S. B., Wang, J., Lu, G.-N., Zhou, P.-G., Hou, S. S., and Xu, S.-N. (2010). GIS-based logistic regression for land slide susceptibility mapping of the Zhongxian segment in the three Gorges area China. *Geomorphol. (Amst).* 115 (1-2), 23–31. doi:10.1016/j.geomorph.2009.09.025

Barrett, C. B., Reardon, T., and Webb, P. (2001). Nonfarm income diversification and household livelihood strategies in rural Africa: Concepts, dynamics and policy implications. *Food Policy* 26 (4), 315–331. doi:10.1016/s0306-9192(01)00014-8

Benz, A. (2016). Framing Modernization interventions: Reassessing the role of migration and Translocality in sustainable mountain development in Gilgit-Baltistan, Pakistan. *Mt. Res. Dev.* 36, 141–152. doi:10.1659/mrd-journal-d-15-00055.1

Bhandari, P. B. (2013). Rural livelihood change? Household capital, community resources and livelihood transition. *J. Rural Stud.* 32, 126–136. doi:10.1016/j.jrurstud.2013.05.001

Binswanger-Mkhize, H. P. (2013). The stunted structural transformation of the Indian economy. *Econ. Political Wkly.* 48 (26-27), 5–13.

Bruce, A, B. (2019). Farm entry and persistence: Three pathways into alternative agriculture in southern Ohio. *J. Rural Stud.* 69, 30–40. doi:10.1016/j.jrurstud.2019.04.007

Cai, F., and Wang, M. (2010). Growth and structural changes in employment in transition China. *J. Comp. Econ.* 38 (1), 71–81. doi:10.1016/j.jce.2009.10.006

Cain, Mead T. (1977). The economic activities of Children in a village in Bangladesh. *Popul. Dev. Rev.* 3 (3), 201–227. doi:10.2307/1971889

Calo, A. (2018). How knowledge deficit interventions fail to resolve beginning farmer challenges. *Agric. Hum. Values* 35 (2), 367–381. doi:10.1007/s10460-017-9832-6

Calus, M., Van Huylenbroeck, G., and Van Lierde, D. (2008). The relationship between farm succession and farm assets on Belgian farms. *Sociol. Rural.* 48 (1), 38–56. doi:10.1111/j.1467-9523.2008.00448.x

Caskie, P., Davis, J., Campbell, D., and Wallace, M. (2002). *An economic study of Farmer early retirement and new entrant Schemes for Northern Ireland*. United Kingdom, London: Queen's University Belfast.

Chitrakar, P. L. (1990). *Planning, agriculture and Farmers: Strategies for Nepal*. Kathmandu, Nepal: Ganesh Devi Chitrakar.

Conkin, P. (2008). *A Revolution down on the Farm: The transformation of American agriculture since 1929*. Kentucky, United States: University Press of Kentucky.

Davidova, S., and Thomson, K. (2013). Family farming: A Europe and Central Asia perspective. Available at https://kar.kent.ac.uk/id/eprint/64951 (Accessed November 16, 2021).

Deshpande, R. S., and Prabhu, N. (2005). Farmers' distress: Proof beyond question. *Econ. Political Wkly.* 40, 4663–4665.

Eifert, B., and Ramachandran, V. (2004). *Competitiveness and private sector development in Africa: Cross-country evidence from the World Bank's investment climate data*. Ithaca, N.Y. Institute for African Development, Cornell University.

Errington, A., and Lobley, M. "Handing over the Reins: A comparative study of intergenerational farm transfers in England, France, Canada and the USA," in 2002 International Congress, Zaragoza, Spain, August 2002 (Agricultural Economics Society Conference).

Filmer, D., and Pritchett, L. (1997). *Environmental degradation and the demand for Children: Searching for the Vicious Circle (World Bank policy research Paper No. 1623)*. Washington, DC: The World Bank.

Food and Agriculture Organization (2021). *Food and agriculture Organization-FAO statistical Yearbook*. Rome, Italy: FAO.

Frost, J. (2017). *Becoming (and Remaining) a Farmer is Hard*. Boston, USA: Weekly Civil Eats Newsletter.

Gale, H. (2003). Age-specific Patterns of exit and entry in U.S. Farming, 1978-1997. *Rev. Agric. Econ.* 25 (1), 168–186. doi:10.1111/1467-9353.00052

Gillespie, G. W., and Johnson, S. E. (2010). Success in farm start-ups in the Northeastern United States. *J. Agric. Food Syst. Community Dev.* 1 (1), 31–48. doi:10.5304/jafscd.2010.011.008

Gioli, G., Khan, T., Bisht, S., and Scheffran, J. (2014). Migration as an adaptation strategy and its gendered implications: A case study from the Upper Indus basin. *Mt. Res. Dev.* 34 (3), 255–265. doi:10.1659/mrd-journal-d-13-00089.1

Glauben, T., Tietje, H., and Weiss, C. (2006). Agriculture on the move: Exploring regional differences in farm exit rates in western Germany. *Jahrb. für Regionalwiss.* 26 (1), 103–118. doi:10.1007/s10037-004-0062-1

Gruère, G., and Sengupta, D. (2011). Bt Cotton and farmer suicides in India: An evidence-based Assessment. *J. Dev. Stud.* 47, 316–337. doi:10.1080/00220388.2010.492863

Gurung, O. P. (1987). *Inter-relationship among pasture, animal Husbandry and agriculture: Case study of Tara*. Winrock: Kathmandu, Nepal: National Resource Management Paper Series No. 2.

Haggblade, S., Hazell, P., and Reardon, T. (2010). The rural non-farm economy: Prospects for growth and poverty Reduction. *World Dev.* 38 (10), 1429–1441. doi:10.1016/j.worlddev.2009.06.008

Heckman, J., Ichimura, H., Smith, J., and Todd, P. (1998). Characterizing selection bias using experimental data. *Econometrica* 66 (5), 1017–1098. doi:10.2307/2999630

Heckman, J., Ichimura, H., and Todd, P. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *Rev. Econ. Stud.* 64 (4), 605–654. doi:10.2307/2971733

Herrero, M., Thornton, P. K., Notenbaert, A. M., Wood, S., Msangi, S., Freeman, H. A., et al. (2010). Smart investments in sustainable food production: Revisiting mixed crop-livestock systems. *Science* 327 (5967), 822–825. doi:10.1126/science.1183725

Huang, Z., Zhang, X., and Zhu, Y. (2008). The role of clustering in rural industrialization: A case study of the footwear industry in Wenzhou. *China Econ. Rev.* 19 (3), 409–420. doi:10.1016/j.chieco.2007.11.001

Hussain, I. (2014). *Urbanization in Pakistan'*. Karachi: Institute of Business Administration.

Imai, K., Gaiha, R., and Garbero, A. (2014). *Poverty reduction during the rural–urban transformation: Rural development is still more important than urbanization (BWPI Working Paper WP 204/2014)*. Manchester: The University of Manchester.

Ingram, J., and Kirwan, J. (2011). Matching new entrants and retiring farmers through farm joint ventures: Insights from the Fresh Start Initiative in Cornwall, UK. *Land Use Policy* 28 (4), 917–927. doi:10.1016/j.landusepol.2011.04.001

International Food Policy Research Institute (2014). *Pakistan rural household panel survey (PRHPS) 2012, round 1*. Washington, D.C: IFPRI.

International Food Policy Research Institute (2015). *Pakistan rural household panel survey (PRHPS) 2012, round 1.5*. Washington, D.C: IFPRI.

International Food Policy Research Institute (2016). *Pakistan rural household panel survey (PRHPS) 2013, round 2*. Washington, D.C: IFPRI.

International Food Policy Research Institute (2017). *Pakistan rural household panel survey (PRHPS) 2014, round 3*. Washington, D.C: IFPRI.

Jan, I., Khattak, M. K., Khan, M. S., Hayat, S., and Rahim, T. (2012). Factors affecting rural livelihood choices in Northwest Pakistan. *Sarhad J. Agric.* 28 (4), 681–688.

Janvry de, A. (1994). Farm-nonfarm synergies in Africa: Discussion. *Am. J. Agric. Econ.* 76 (5), 1183–1185. doi:10.2307/1243414

Karan, P. P., and Ishii, H. (1995). *Nepal: A Himalayan Kingdom in transition*. Tokyo, New York, Paris: United Nations University Press.

Kimhi, A., and Bollman, R. (1999). Family farm dynamics in Canada and Israel: The case of farm exits. *Agric. Econ.* 21 (1), 69–79. doi:10.1016/s0169-5150(99)00015-8

Knight, J., Deng, Q., and Li, S. (2011). The puzzle of migrant labour shortage and rural labour surplus in China. *China Econ. Rev.* 22 (4), 585–600. doi:10.1016/j.chieco.2011.01.006

Kumar, S. K., and Hotchkiss, D. (1988). *Agricultural production, and Nutrition in Hill areas of Nepal (research report No. 69)*. Washington, DC: International Food Policy Research Institute.Consequences of Deforestation for Women's time allocation

Lobao, L., and Meyer, K. (2001). The great agricultural transition: Crisis, change, and social consequences of twentieth century US farming. *Annu. Rev. Sociol.* 27 (1), 103–124. doi:10.1146/annurev.soc.27.1.103

Loening, J., and Mikael, I. (2009). *Ethiopia: Diversifying the rural economy an Assessment of the investment climate for small and informal enterprises*. World Bank Economic Sector Work. Vol. 49564-ET.

Lusher Shute, L. (2011). *Building a Future with Farmers: Challenges Faced by Young, American Farmers and a national strategy to Help them succeed*. New York: National Young farmers Coalition.

Mishra, A. K., and El-Osta, H. S. (2016). Determinants of decisions to enter the U.S. farming sector. *J. Agric. Appl. Econ.* 48 (1), 73–98. doi:10.1017/aae.2015.25

Mishra, S. (2006). Farmers' suicides in Maharashtra. *Econ. Political Wkly.* 41, 1538–1545.

Pakistan Bureau of Statistics (2018). *Agricultural statistics of Pakistan*. Islamabad, Pakistan: Government of Pakistan, Statistics Division.

Pakistan Bureau of Statistics (2019). *Agriculture statistics of Pakistan*. Islamabad, Pakistan: Government of Pakistan, Statistics Division.

Pietola, K., Vare, M., and Lansink, A. O. "Farmer's exit decisions and early retirement Programs in Finland," in 2002 International Congress, Zaragoza, Spain, August, 2002 (European Association of Agricultural Economists).

Rizwan, M., Deyi, Z., Nazir, A., Ostai, R., Traore, L., and Sargani, G. R. (2017). Determinants and choices of off-farm work among Rice farmers in a developing country. *J. Animal Plant Sci.* 27 (6), 1993–2002.

Rose, E. (2001). Ex ante and ex post labor supply response to risk in a low-income area. *J. Dev. Econ.* 64 (2), 371–388. doi:10.1016/s0304-3878(00)00142-5

Ruan, J., and Zhang, X. (2009). Finance and cluster-based industrial development in China. *Econ. Dev. Cult. Change* 58 (1), 143–164. doi:10.1086/605208

Shahzad, M. A., Abubakr, S., and Fischer, C. (2021a). Factors affecting farm succession and occupational choices of Nominated farm successors in Gilgit-Baltistan, Pakistan. *Agriculture* 11 (12), 1203. doi:10.3390/agriculture11121203

Shahzad, M. A., Ahmed, V., and Fischer, C. (2021b). Status and determinants of other gainful activities by farmers in mountainous rural regions of Gilgit-Baltistan, Pakistan. *J. Mt. Sci.* 18, 2520–2539. doi:10.1007/s11629-021-6673-y

Sianesi, B. (2004). An evaluation of the Swedish system of active labor market Programs in the 1990s. *Rev. Econ. Statistics* 86 (1), 133–155. doi:10.1162/003465304323023723

State Bank of Pakistan (2015). *Annual report 2014-2015 (state of the economy)*. Pakistan, Karachi: State Bank of Pakistan.

Tahir, M., Khan, H., and Qahar, A. (2012). Factors determining off-farm employment on small farms Iin district Nowshehra of North west Pakistan. *Sarhad J. Agric.* 28 (2), 333–344.

Vare, M., and Heshmati, A. (2004). *Perspectives on the early retirement decisions of Farming Couples*. Bonn, Germany: Institute for the Study of Labor. Discussion Paper No. 1342.

Wang, X., Huang, J. K., Zhang, L. X., and Rozelle, S. (2011). The rise of migration and the fall of self employment in rural China's labor market. *China Econ. Rev.* 22 (4), 573–584. doi:10.1016/j.chieco.2011.07.005

Whitehead, I., and Millard, N. (2000). *Contemporary issues of county farm estates in England and Wales*. London, United Kingdom: Royal Institute of Chartered Surveyors.

Williams, F., and Farrington, J. "Succession and the future of farming: Problem or perception?," in Conference on the Rural Citizen: Governance, Culture and Wellbeing in the 21st Century, Plymouth, UK, 2006 (University of Plymouth).

Wooldridge, J. M. (2009). *Introductory Econometrics: A modern approach*. OH, USA: South-Western Cengage Learning: Mason.

World Bank (2007). *China's modernizing labor market: Trends and emerging challenges*. Washington, D.C., United States: World Bank. Synthesis Report for the ESW Component of the China Labor Market AAA Program.

World Bank (2008). *World development report 2008: Agriculture for development*. Washington, D.C., United States: World Bank Publications.

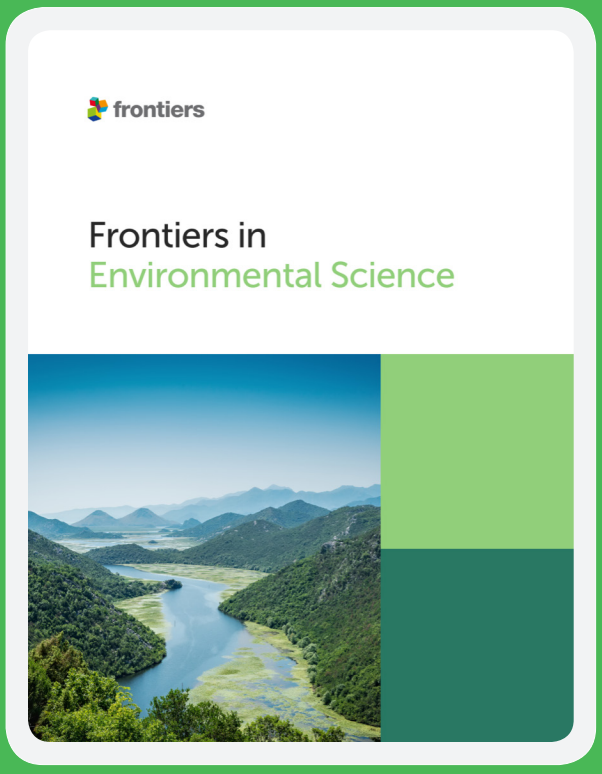
Frontiers in
Environmental Science

Explores the anthropogenic impact on our natural world

An innovative journal that advances knowledge of the natural world and its intersections with human society. It supports the formulation of policies that lead to a more inhabitable and sustainable world.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact