Research Article

Spatial Interpolation of Annual Runoff in Ungauged Basins Based on the Improved Information Diffusion Model Using a Genetic Algorithm

Mei Hong,[1,2] Ren Zhang,[1,2] Dong Wang,[3] Longxia Qian,[1] and Zhenghua Hu[2]

[1]*Research Center of Ocean Environment Numerical Simulation, Institute of Meteorology and Oceanography,*
 PLA University of Science and Technology, Nanjing, China
[2]*Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disaster,*
 Nanjing University of Information Science & Technology, Nanjing 210044, China
[3]*Key Laboratory of Surficial Geochemistry, Ministry of Education, Department of Hydrosciences, School of Earth Sciences and*
 Engineering, State Key Laboratory of Pollution Control and Resource Reuse, Nanjing University, Nanjing 210093, China

Correspondence should be addressed to Ren Zhang; majin_rene@sina.com and Dong Wang; wangdong@nju.edu.cn

Prediction in Ungauged Basins (PUB) is an important task for water resources planning and management and remains a fundamental challenge for the hydrological community. In recent years, geostatistical methods have proven valuable for estimating hydrological variables in ungauged catchments. However, four major problems restrict the development of geostatistical methods. We established a new information diffusion model based on genetic algorithm (GIDM) for spatial interpolating of runoff in the ungauged basins. Genetic algorithms (GA) are used to generate high-quality solutions to optimization and search problems. So, using GA, the parameter of optimal window width can be obtained. To test our new method, seven experiments for the annual runoff interpolation based on GIDM at 17 stations on the mainstream and tributaries of the Yellow River are carried out and compared with the inverse distance weighting (IDW) method, Cokriging (COK) method, and conventional IDMs using the same sparse observed data. The seven experiments all show that the GIDM method can solve four problems of the previous geostatistical methods to some extent and obtains best accuracy among four different models. The key problems of the PUB research are the lack of observation data and the difficulties in information extraction. So the GIDM is a new and useful tool to solve the Prediction in Ungauged Basins (PUB) problem and to improve the water management.

1. Introduction

Prediction in Ungauged Basins (PUB) [1] is an important task for water resources planning and management and remains a fundamental challenge for the hydrological community. Prediction in Ungauged Basins (PUB) was identified as a key issue in hydrological studies by IAHS. Accurate estimates of hydrologic variables at ungauged sites such as streamflow allow objective, quantitative, and statistical decision-making with respect to water resources management and natural hazard assessments. The lack of data for model calibration and verification in ungauged basins requires the hydrological regionalization [2] to transfer information (e.g., model parameters) from gauged catchments. The regionalization allows estimating parameter values of hydrological predictive tools without calibration. Regionalization can be defined as the transfer of information from one catchment to another [2]. This transfer is typically from gauged to ungauged catchments (e.g., [3, 4]). Its aim is to estimate parameter values of hydrological models for any/every grid cell, subcatchment, or large geographic region without a need for calibration or "tune" of the model to get the best fit.

Over the years, regionalization has received increasing attention from the hydrological community. A number of regional models are currently available, including (1) proxy-basin method [5, 6], (2) spatial interpolation method, for

instance, linear interpolation by Guo et al. [7], the inverse distance weighting (IDW) interpolation by Di Piazza et al. (2001), and Kriging interpolation by Vandewiele and Elias [8], (3) clustering approach [9, 10], (4) bi- and multivariate regression method [11, 12], and (5) one step regression–regional calibration [13]. Among those, spatial interpolation method is one of the earliest and most widely used methods, which estimates the value of unknown spatial data based on known spatial data. Its essence is the spatial forecasting of the whole unknown region using a few known points. Deterministic and geostatistical techniques are two main groupings of spatial interpolation techniques to produce a continuous surface from point measurements. Deterministic interpolation techniques create surfaces from measured points using mathematical functions, which are based on either the extent of similarity (e.g., inverse distance weighted) or the degree of smoothing (e.g., radial basis functions). Geostatistical interpolation techniques (e.g., Kriging) utilize both the mathematical and the statistical properties of the measured points [14, 15]. Recent studies highlight that geostatistical interpolation, which has been originally developed for the spatial interpolation of point data (see e.g., [16]), can be effectively applied to the prediction problem of the streamflow regime in ungauged basins [17, 18].

Recently, geostatistical methods have proven valuable for estimating hydrological variables in ungauged catchments. In all geostatistical methods, the traditional Kriging and its related algorithms (e.g., Universal Kriging or Cokriging (COK)) are the most widely used. Skøien and Blöschl [18], for instance, developed the topological Kriging technique (or top-Kriging), which accounts for hydrodynamic and geographical dispersion. Their results indicate that this technique can not only outperform deterministic runoff models in regions where stream gauge density is sufficiently high, as it avoids problems with input data errors and parameter identifiability, but also provide more robust estimates than regional regression models [19]. Comparison of top-Kriging with Physiographical-Space Based Interpolation (PSBI) highlights the complementary utility of the two methods for headwater and larger scale catchments [20].

However, four major limitations of geostatistical methods (e.g., Kriging and its related algorithms) are presented.

(1) The ordinary Kriging is defined as a "best linear unbiased estimator." Kriging is "linear" because its estimates are calculated by a linear equation. While the change of runoff is a nonlinear process, these will cause some deviations.

(2) The methods need more sites for modelling, generally more than eight sites [21]. So it is more feasible when it is used in the large watershed interpolating. When the study sites are little in minor watershed, the method can do nothing for the space structure of hydrology variables.

(3) The interpolation process of many geostatistical methods needs to add water balance constraints, which controls the export flows of each subbasin. Before the runoff calculation, the runoff data in the study should be normalized [22]. In the interpolation

process, in order to measure the correlation between the basins, the distance between the subbasins of the interpolation algorithm need to be redefined [23]. The calculation of the entire process is complex and the conversion is troublesome.

(4) The application of these methods is mainly in European natural basin [24]. The application in the watershed acutely impacted by human activities needs to be verified.

In order to overcome the above four questions, we need to find new spatial interpolation method which is more effective and rational. So we introduce the information diffusion model (IDM) in our paper. IDM is a useful method to deal with the small sample problem [25, 26]. Spreading the observed data can extract many additional information based on the diffusion methods. Huang [25] can easily determine simple window width (SWW) with incomplete data based on the nearby criteria. This method was widely used for reliability of risk assessment [27–29]. But IDM with SWW (SIDM) cannot accurately calculate the hydrological or meteorological data which follow abnormal distribution. To solve the problem, Wang et al. [30] presented the optimal window width based on IDM (OIDM) using the principle of least mean squared errors. But the optimal window width (OWW) may easily cause the local optimal problem. Genetic Algorithm (GA) is a metaheuristic inspired by the process of natural selection that belongs to the larger class of evolutionary algorithms (EA). Genetic algorithms are commonly used to generate high-quality solutions to optimization and search problems [31]. Hence, to get the global optima diffusion coefficients, Bai et al. [32] propose a new information diffusion model using a GA (GIDM) to interpolate the river runoff. But they just used GIDM method to interpolate the river runoff time series. Now our paper expand the idea of GIDM and use them to spatially interpolate the river runoff in ungauged basins.

No previous hydrological literature reported that the IDM has been used for establishing a hydrological model for hydrological spatial interpolation. So we explain the principle of the SWW and OWW in Section 2; we also discuss the new method how to improve the information diffusion model based on GA. In Section 3, we illustrate interpolation of river runoff based on IDM in detail. To test our new method, seven experiments for the runoff interpolation based on the GIDM in the Yellow River are carried out in Section 4, compared with SIDM, OIDM, IDW, and COK. Finally, Section 5 summarizes and discusses the results.

2. Information Diffusion Model and the Window Width

2.1. Simple Window Width and Optimal Window Width. The principle of information diffusion and simple window width (SWW) is discussed in the Appendix and was introduced in the previous literature [25].

The SWW method extracted more data from the small sample [26, 33]. However, when the population does not follow a normal distribution, the method is invalid. Based on

the idea of the least mean square error, the optimal window width (OWW) d can be obtained as follows [30]:

$$d^0 = \frac{\max(X) - \min(X)}{n-1},$$

$$d_i^{j+1}$$

$$= \left\{ \frac{\widehat{f}^j(x_i)(d_j)^4}{2n\sqrt{\pi}\left[\widehat{f}^j(x_i + d_i) - 2\widehat{f}^j(x_i) + \widehat{f}^j(x_i - d_i)\right]^2} \right\}^{0.2},$$

$$(i = 1, \dots, n),$$

$$d^{j+1} = \text{mean}\left(d_i^{j+1}\right),$$

$$\widehat{f}^{j+1}(x_i) = \frac{1}{\sqrt{2\pi}nd^j}\sum_j^k \mu\left[\frac{x - x_i}{d^j}\right],$$

where i and j mean the ordinal number of records and iterations; initial iterative value is d^0.

$$d = \frac{1}{k - i + 1}\sum_{j=i}^k d^j. \tag{2}$$

The OWW based on the mean value will cause the local optimization problem [31]. So GA is introduced.

2.2. Searching the General Optimal Window Width Based on GA. Based on the principle of natural genetics and biological evolution, Genetic Algorithms (GA) can effectively avoid the "local optima" problem [31]. So we use GA to search the global optimal diffusion coefficients. For window width searching, the combination of the IDM and GA includes three major phases. Firstly, the GA initializes a population that compounds random codes from the search domain $(0, (b - a)/3]$ [30], where a and b are the minimum and maximum value of the samples. Then we carry out the evaluation of the fitness of all chromosomes. Based on Wang et al. [30], the window width d can be obtained as follows:

$$d_i^5 = \frac{\widehat{f}(x_i)}{2n\sqrt{\pi}\left[\widehat{f}''(x_i)\right]^2}, \tag{3}$$

where x_i means different records from sample $(i = 1, 2, \dots, n)$ and $\widehat{f}(x)$ denotes the information diffusion estimator. The second-order schemes are motivated by

$$\widehat{f}''(x_i) = \frac{\widehat{f}(x_i + d_i) - 2\widehat{f}(x_i) + \widehat{f}(x_i - d_i)}{d_i^2}, \tag{4}$$

where

$$d_i = \frac{\widehat{f}(x_i)}{2n\sqrt{\pi}\left[\widehat{f}(x_i + d_i) - 2\widehat{f}(x_i) + \widehat{f}(x_i - d_i)\right]^2}. \tag{5}$$

The global optimal d can be searched as follows:

CalFitness (d)

$$= \sum_{i=1}^n \left| d - \frac{\widehat{f}(x_i)}{2n\sqrt{\pi}\left[\widehat{f}(x_i + d) - 2\widehat{f}(x_i) + \widehat{f}(x_i - d)\right]^2} \right|. \tag{6}$$

The evolutionary processes can be found in [31]. So we can get the improved window width (IWW).

The above sentences explain the techniques of IDM and improved IDM; how to interpolate the runoff based on IDM will be discussed in following section.

3. Information Diffusion Method with Fuzzy Inference for the Runoff Estimation

The IDM based on the numerical method [28] is as follows. Let $(X, Y) = \{(x_i, y_i) \mid i = 1, 2, \dots, n\}$ be a training set of monthly discharge data on $\mathbb{R}^2$ ($\mathbb{R}$ is the real line), where input x means the index of records sorted by meteorological observations or chronological order, like soil moisture, precipitation, evaporation, and so forth, and y means the river discharge.

Let V be the range of y and U be the domain of x. u will denote the element of U, and the same for v by V. Let

$$U = \{u_j, \ j = 1, 2, \dots, t\},$$
$$V = \{v_k, \ k = 1, 2, \dots, l\}. \tag{7}$$

Dealing with membership functions based on (A.1) and (A.2), the following equation can be obtained:

$$\mu_U(u_j) = \frac{1}{td_U\sqrt{2\pi}}\sum_{j=1}^t \exp\left[-\frac{(u - u_j)^2}{2d_U^2}\right], \quad u \in U,$$

$$\mu_V(v_k) = \frac{1}{td_V\sqrt{2\pi}}\sum_{j=1}^t \exp\left[-\frac{(v - v_k)^2}{2d_V^2}\right], \quad v \in V. \tag{8}$$

d_U and d_V are window widths. (u_j, v_k) is called an illustrating point. The information gain of (u_j, v_k) is as follows:

$$q_{jk}(x_i, y_i) = \mu_{U_j}(x_i) \times \mu_{V_k}(y_i). \tag{9}$$

Then,

$$Q_{jk} = \sum_{i=1}^n q_{jk}(x_i, y_i),$$

$$Q = \begin{matrix} & \begin{matrix} v_1 & v_2 & \cdots & v_l \end{matrix} \\ \begin{matrix} u_1 \\ u_2 \\ \vdots \\ u_t \end{matrix} & \begin{pmatrix} Q_{11} & Q_{12} & \cdots & Q_{1l} \\ Q_{21} & Q_{22} & \cdots & Q_{2l} \\ \vdots & \vdots & \vdots & \vdots \\ Q_{t1} & Q_{t2} & \cdots & Q_{tl} \end{pmatrix} \end{matrix}. \tag{10}$$

A fuzzy relation matrix R [34] is as follows:

$$R = \begin{matrix} & \begin{matrix} v_1 & v_2 & \cdots & v_l \end{matrix} \\ \begin{matrix} u_1 \\ u_2 \\ \vdots \\ u_t \end{matrix} & \begin{pmatrix} R_{11} & R_{12} & \cdots & R_{1l} \\ R_{21} & R_{22} & \cdots & R_{2l} \\ \vdots & \vdots & \vdots & \vdots \\ R_{t1} & R_{t2} & \cdots & R_{tl} \end{pmatrix} \end{matrix}, \tag{11}$$

which can be got from an information matrix Q by using

$$R = \{r_{jk}\}_{t \times l} = \{r(u_j, v_k)\}_{t \times l},$$

$$r_{jk} = \frac{Q_{jk}}{s_k}, \tag{12}$$

$$s_k = \max_{1 \le j \le t} Q_{jk}.$$

Then A is denoted as the input fuzzy set to calculate the output fuzzy set B;

$$\mu_A(u_j) = \sum_{z=1}^{t} \frac{q}{u_z}, \quad q = \begin{cases} 1, & z = j \\ 0, & z \ne j. \end{cases} \tag{13}$$

The fuzzy inference formula is used as follows:

$$B = A \circ R, \tag{14}$$

in which max–min fuzzy composition rule is denoted as operator "∘";

$$\mu_B(v_k) = \max_{u_j \in U} \left\{ \min \left[\mu_A(u_j), r(u_j, v_k) \right] \right\}, \quad v_k \in V, \tag{15}$$

where $r(u_j, v_k) \in (0, 1]$; so we can get

$$\mu_B(v_k) = \max_{u_j \in U} \left\{ r(u_j, v_k) \right\}. \tag{16}$$

Finally, we can generate the gravity center of the fuzzy set as the output:

$$\tilde{y}_j = \frac{\sum_{k=1}^{l} v_k \mu_B(v_k)}{\sum_{k=1}^{l} \mu_B(v_k)}. \tag{17}$$

In general, the given sample (X, Y) is used to construct the relationship between river discharge and its meteorological factor or its antecedent values as follows:

$$\tilde{Y} = f(X_i), \tag{18}$$

where X_i is an input vector including $x_1, \ldots, x_i, \ldots, x_m$, and $\tilde{Y}$ means the flow in the ungauged basins. So the value of river runoff in the ungauged basins can be obtained by the IDM method.

4. Case Study

To test the runoff interpolation effectiveness of our model (GIDM), we carried out six experiments. These experiments can divide into five groups. Experiments 1 and 2 are spatial interpolation experiments of runoff on mainstream of the Yellow River; Experiments 3 and 4 are spatial interpolation experiments of runoff on tributaries of the Yellow River. Experiments 5 and 6 are spatial interpolation experiments of runoff on the mainstream and tributaries of the Yellow River. These six experiments are carried out to test the spatial interpolation and prediction ability of GIDM model for the mainstream and tributaries of the same river basin. Finally, experiment 7 is the spatial interpolation experiment of runoff
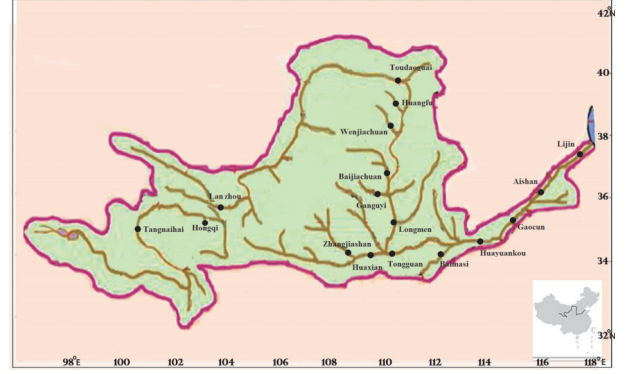

FIGURE 1: Map of the Yellow River basins showing the location of hydrological stations.

in Daying mine region in Guizhou Province (representing nonclosure small watershed of no runoff information), which is used to validate the spatial interpolation and prediction ability of GIDM model for the minor watershed with few hydrological sites. An application of GIDM for runoff interpolation is compared with SIDM, OIDM, IDW, and COK based on the same date.

4.1. Study Area and Data. Tangnaihai, Lanzhou, Toudaoguai, Longmen, Tongguan, Huayuankou, Gaocun, Aishan, and Lijin stations on mainstream of the Yellow River and Hongqi, Huangfu, Wenjiachuan, Baijiachuan, Fanguyi, Zhangjiashan, Baimasi, and Huaxian stations on tributaries of the Yellow River have been selected for the experiments 1–6 (see Figure 1). The Yellow River is the third-longest river in Asia, following the Yangtze River and Yenisei River, and the sixth-longest in the world at the estimated length of 5,464 km. Originating in the Bayan Har Mountains in Qinghai province of western China, it flows through nine provinces, and it empties into the Bohai Sea near the city of Dongying in Shandong province. The Yellow river basin has an east–west extent of about 1,900 kilometers and a north–south extent of about 1,100 km. Its total basin area is about 742,443 square kilometers. So the hydrological stations on the Yellow River are selected for its importance to test the performance of our model.

Yu sha, Baina, Gaofeng, Xiangshui, and Liulong stations in Daying mine region in Guizhou Province (representing nonclosure small watershed of no runoff information) have been selected for the experiment 7.

4.2. Experiments on Estimating Annual Runoff Time Series at Stations on Mainstream of the Yellow River. In this section, two real examples of the annual runoff data ($\times 10^8$ m^3) from 2002 to 2011, taken at nine stations on mainstream of the Yellow River (Figure 1), are presented on mainstream of the same river basin by different models.

4.2.1. Experiment 1: Interpolating Annual Runoff of Two Stations Based on the Data of Other Seven Stations. The sites are selected as Tangnaihai, Lanzhou, Toudaoguai, Longmen, Tongguan, Huayuankou, Gaocun, Aishan, and Lijin stations

TABLE 1: The latitude, longitude, altitude, catchment area, and the annual runoff from 2002 to 2011 of nine sites on mainstream of the Yellow River.

Site	Tangnaihai	Lanzhou	Toudaoguai	Longmen	Tongguan	Huayuankou	Gaocun	Aishan	Lijin
Latitude (°N)	35.6	36	40.1	35.7	34.5	34.8	35.1	36.3	35.6
Longitude (°E)	100.1	103.8	111.1	110.6	110.3	114.1	115.1	116.2	111
Altitude (m)	3320	1520	1620	421	327.94	205	78	197.2	8
Catchment area (km^2)	121972	222551	367898	497552	682141	730036	734146	749136	751869
The runoff in 2011 ($\times 10^8$ m^3)	211.2	284.1	162.9	169.4	259.6	287.1	262.3	238.6	184.2
The runoff in 2010 ($\times 10^8$ m^3)	197.1	314	191.2	207.3	262.5	276.3	258.3	242	193
The runoff in 2009 ($\times 10^8$ m^3)	263.5	304.7	169.6	178.3	206.5	232.2	208.9	187.9	132.9
The runoff in 2008 ($\times 10^8$ m^3)	174.6	285.1	164.1	177.6	204.8	236.1	220.8	197.1	145.6
The runoff in 2007 ($\times 10^8$ m^3)	189.0	307.0	189.3	205.9	250.4	269.7	259.8	248.7	204.0
The runoff in 2006 ($\times 10^8$ m^3)	141.3	298.9	174.9	199.6	233.4	281.1	265.9	246.6	191.7
The runoff in 2005 ($\times 10^8$ m^3)	255.0	291.1	150.2	169.2	230.8	257.0	243.4	245.4	206.8
The runoff in 2004 ($\times 10^8$ m^3)	151.3	238.5	127.6	158.5	197.3	240.5	231.0	233.7	198.8
The runoff in 2003 ($\times 10^8$ m^3)	171.6	219.7	115.6	162.3	261.4	272.7	257.6	240.0	192.6
The runoff in 2002 ($\times 10^8$ m^3)	105.8	235.8	122.8	156.6	174.7	195.6	157.7	107.9	41.90

TABLE 2: The corresponding diffusion coefficients of Experiment 1.

	SWW	OWW	IWW
X_1	5.3742	14.2077	0.6941
X_2	6.7812	15.4218	0.7732
X_3	9.8157	22.4517	1.2850
X_4	4.8641	17.3608	0.4585
Y	5.5214	32.9110	0.3155

on mainstream of the Yellow River (Figure 1). As Table 1 shown, the latitude, longitude, altitude, and catchment area of nine sites are known. The natural annual runoff data of seven sites (Tangnaihai, Toudaoguai, Longmen, Tongguan, Huayuankou, Gaocun, and Aishan) are known for modelling, while the natural annual runoff data of Lanzhou site (located upstream of the Yellow River) and Lijin site (located downstream of the Yellow River) are seen as unknown for testing, which are obtained through spatial interpolation. The actual annual runoff data of Lanzhou site and Lijin site in Table 1 can be used to test the results of spatial interpolation.

The specific procedure of the test is as follows.

Step 1. The data of seven sites (Tangnaihai, Toudaoguai, Longmen, Tongguan, Huayuankou, Gaocun, and Aishan) on mainstream of the Yellow River is used as the modelling sample, in which latitude, longitude, altitude, and catchment area are seen as input data and the natural runoff data in 2011 are seen as output data. So the input data is $X = (x_1, x_2, x_3, x_4) = \{(35.6, 100.1, 3320, 121972), (40.1, 111.1, 1620, 367898), \ldots, (36.3, 116.2, 197.2, 749136)\}$ and the output data is $Y = \{211.2, 284.1, 162.9, 169.4, \ldots, 184.2\}$.

Step 2. Calculate the window width of IDM based on the discussion in the Section 2. Table 2 is the results of SWW, IWW, and OWW.

Step 3. Appropriate illustrating points are very important because they present the basic information of the system. The

illustrating space can be reconstructed based on the statistical analysis.

Step 4. The input data of two test sites Lanzhou and Lijin is $X = (x_1, x_2, x_3, x_4) = \{(36, 103.8, 1520, 222551), (35.6, 111, 8, 751869)\}$, respectively, calculating the final output (the natural runoff of two sites in 2011) using (8) to (18).

Step 5. Similar as the above calculation progress of the natural runoff of two sites in 2011, we can also get the annual natural runoff results of two sites from 2002 to 2010.

To further test the interpolating results of our GIDM model, we not only use the SIDM and OIDM methods to be compared with the GIDM model, but also choose the IDW and COK interpolation method, which are two most commonly spatial interpolation methods now, to be compared with the GIDM model. Specific principles and operational steps of the IDW and COK interpolation method can be seen in the related literature [35–37], which are not described in detail here.

Figure 2 shows interpolated and observed runoffs in Lanzhou station and Lijin station. The preformation of GIDM is better than the other four models for two sites. From Figure 2(a), some obvious oversimulations exist in 2003, 2004, and 2006, which are the same as undersimulations in 2009, 2010, and 2011 obtained by SIDM, OIDM, IDW, and COK, while the GIDM exhibits a good correlation. Although a few discrepancies between simulated data and observed data using GIDM exist (such as from 2004 to 2006), considering the limited samples, the general tendency can be accepted.

The Mean Absolute Percentage Error (MAPE, [38]), the Root Mean Square Error (RMSE, [39, 40]), the Nash and Sutcliffe coefficient (*E*, [41]), and the coefficient of correlation (*R*, [39]) are selected to test the model. The RMSE, *R*, *E*, and MAPE of different models are shown in Table 3. From Table 3, the IDW performs worst and GIDM performs best among

TABLE 3: The RMSE, R, E, and MAPE of five different models in Lanzhou site and Lijin site (where boldface font indicates the best performance).

| Site	Model	RMSE	R	E	MAPE
	IDW	12.39	0.8571	0.7426	41.36%
	COK	11.32	0.8897	0.7918	35.71%
Lanzhou	SIDM	10.72	0.9016	0.8127	32.65%
	OIDM	10.19	0.9089	0.8194	29.71%
	GIDM	**6.48**	**0.9367**	**0.8457**	**12.33%**
	IDW	10.26	0.7983	0.6860	52.73%
	COK	9.07	0.8258	0.7169	43.34%
Lijin	SIDM	9.01	0.8617	0.7607	41.28%
	OIDM	8.25	0.8865	0.7815	36.45%
	GIDM	**5.97**	**0.9260**	**0.8213**	**18.42%**
	IDW	11.325	0.8277	0.7143	47.05%
	COK	10.195	0.85775	0.75435	39.53%
The average	SIDM	9.865	0.88165	0.7867	36.97%
	OIDM	9.22	0.8977	0.80045	33.08%
	GIDM	**6.225**	**0.9314**	**0.8335**	**15.38%**
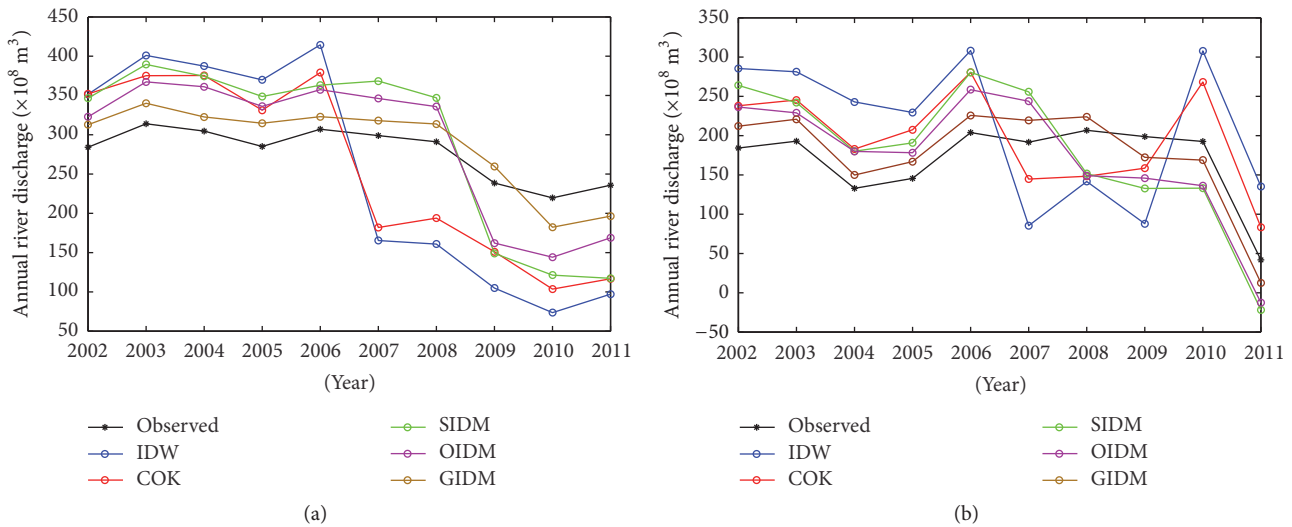
FIGURE 2: Comparison of observed runoff and interpolated runoff forced by IDW, COK, SIDM, OIDM, and GIDM at gauges Lanzhou (a) and Lijin (b) from 2002 to 2011.

five different models. For example, in Lanzhou site from 2002 to 2011, considering a high value of $169.1500 \times 10^8 \, \text{m}^3$ (the average annual runoff) at the Lanzhou gauging station, the GIDM with an RMSE value of $6.48 \times 10^8 \, \text{m}^3$ performed satisfactorily up to the interpolation. Moreover, the GIDM obtained the best R, E, and MAPE statistics of 0.9367, 0.8457, and 12.33% in Lanzhou site, respectively, while the MAPEs of other four models are all over 28%. So the interpolation results of GIDM not only are better than those of the traditional IDW and COK interpolation method, but also can improve traditional IDM interpolation by about 10–40%. Therefore, based on Table 3, the averages of MAPE, RMSE, E, and R of GIDM are 15.38%, 6.225 ($\times 10^8 \, \text{m}^3$), 0.8335, and 0.9314, so the GIDM has consistency and reliable robustness.

The test site Lanzhou is located in the upper reaches of the Yellow River, which is little affected by human activities, while the test site Lijin is located in the lower reaches of the Yellow River, which is intensely influenced by human activities. The site observation data implies the impact of human activities to the natural runoff. Under this case, in this study area the GIDM method is used to research the spatial specificity of runoff and the RMSE, R, E, and MAPE statistics of interpolation results is 5.97 ($\times 10^8 \, \text{m}^3$), 0.9260, 0.8213, and 18.42%, which is a little worse than those of Lanzhou site. It indicates that the interpolation and prediction in the region intensely influenced by human activities are more difficult than those in the region little affected by human activities, but on the whole its interpolation results are still accurate,

TABLE 4: The RMSE, *R*, *E*, and MAPE of five different models in Lanzhou site, Huayuankou site, and Lijin site (where boldface font indicates the best performance).

Site	Model	RMSE	R	E	MAPE
Lanzhou	IDW	14.26	0.8324	0.7269	47.59%
	COK	13.14	0.8496	0.7348	43.88%
	SIDM	11.86	0.8807	0.7715	39.39%
	OIDM	10.94	0.8914	0.7952	35.28%
	GIDM	**7.93**	**0.9183**	**0.8274**	**18.45**%
Huayuankou	IDW	14.41	0.8376	0.7213	45.60%
	COK	12.31	0.8783	0.7762	38.15%
	SIDM	12.35	0.8718	0.7723	38.63%
	OIDM	10.84	0.9008	0.8016	34.18%
	GIDM	**8.07**	**0.9243**	**0.8317**	**26.67**%
Lijin	IDW	11.17	0.7916	0.6817	54.54%
	COK	11.45	0.8341	0.7258	54.89%
	SIDM	9.6	0.8516	0.742	43.63%
	OIDM	7.31	0.9083	0.8062	25.38%
	GIDM	**11.17**	**0.7916**	**0.6817**	**54.54**%
The average	IDW	14.13	0.8072	0.7015	51.19%
	COK	12.21	0.8398	0.7309	44.19%
	SIDM	11.89	0.8622	0.7565	44.97%
	OIDM	10.46	0.8813	0.7796	37.36%
	GIDM	**7.77**	**0.9170**	**0.8218**	**20.50**%

which indicates that the GIDM method can reflect changes in the underlying surface to a certain extent and ensure the reliability of the results of runoff interpolation.

4.2.2. Experiment 2: Interpolating Annual Runoff of Three Stations Using the Data of Other Six Stations. In experiment 2, from 2002 to 2011, the annual natural runoff data of six sites (Tangnaihai, Toudaoguai, Longmen, Tongguan, Gaocun, and Aishan) in Table 1 are known for modelling, while the annual natural runoff data of Lanzhou site (located upstream of the Yellow River), Huayuankou site (located midstream of the Yellow River), and Lijin site (located downstream of the Yellow River) are seen as unknown for testing (Figure 1), which can be obtained through spatial interpolation. The actual annual runoff data of Lanzhou, Huayuankou, and Lijin site in Table 1 can be used to test the results of spatial interpolation. Through similar calculation steps as experiment 1, the annual natural runoff of three sites from 2002 to 2011 can be obtained.

Figure 3 describes reconstructed and observed runoff from different models. The correlation of the GIDM between reconstructed and observed runoff is best among the five different models, although some slight oversimulations and undersimulations exist. Table 4 shows a comparison among different five models and the GIDM method obtains best accuracy among five different models in terms of different evaluation measures. The average of RMSE, *R*, *E*, and MAPE of the GIDM is 7.77 ($\times 10^8$ m³), 0.9170, 0.8218, and 20.50%. Comparing with the average results of GIDM in Table 3, we can see that the interpolating results are worse in experiment 2 than those in experiment 1. In addition,

simulations of all five models gradually become bad based on more data for training. That is because in experiment 2 there are only 6 sites to train and the samples contain less information of the runoff than that in experiment 1. But the average MAPE of GIDM is 20.50% and the average *R* is above 0.9, which indicates that the general results of GIDM are still good and could be acceptable, considering the only six training samples.

Similar as the experiment 1, in this experiment, the interpolation and prediction in the region intensely influenced by human activities (Huayuankou and Lijin sites) are more difficult than those in the region little affected by human activities (Lanzhou site), but on the whole their interpolation results are still accurate.

4.3. Experiments on Estimating Annual Runoff Time Series at Stations on the Tributaries of the Yellow River. In this section, two real examples of the annual runoff data ($\times 10^8$ m³) from 2002 to 2011, taken at eight stations on the tributaries of the Yellow River (Figure 1) are shown to illustrate the implementation on the tributaries of the same watershed by different models.

4.3.1. Experiment 3: Interpolating Annual Runoff of Two Stations Using the Data of Other Six Stations. In experiment 3, from 2002 to 2011, the annual natural runoff data of six sites (Huangfu, Wenjiachuan, Baijiachuan, Ganguyi, Zhangjiashan, and Huaxian) on the tributaries of the Yellow River are known for modelling, while the annual natural runoff data of two sites (Hongqi and Baimasi) on the tributaries of the Yellow River are seen as unknown for testing (Figure 1), which
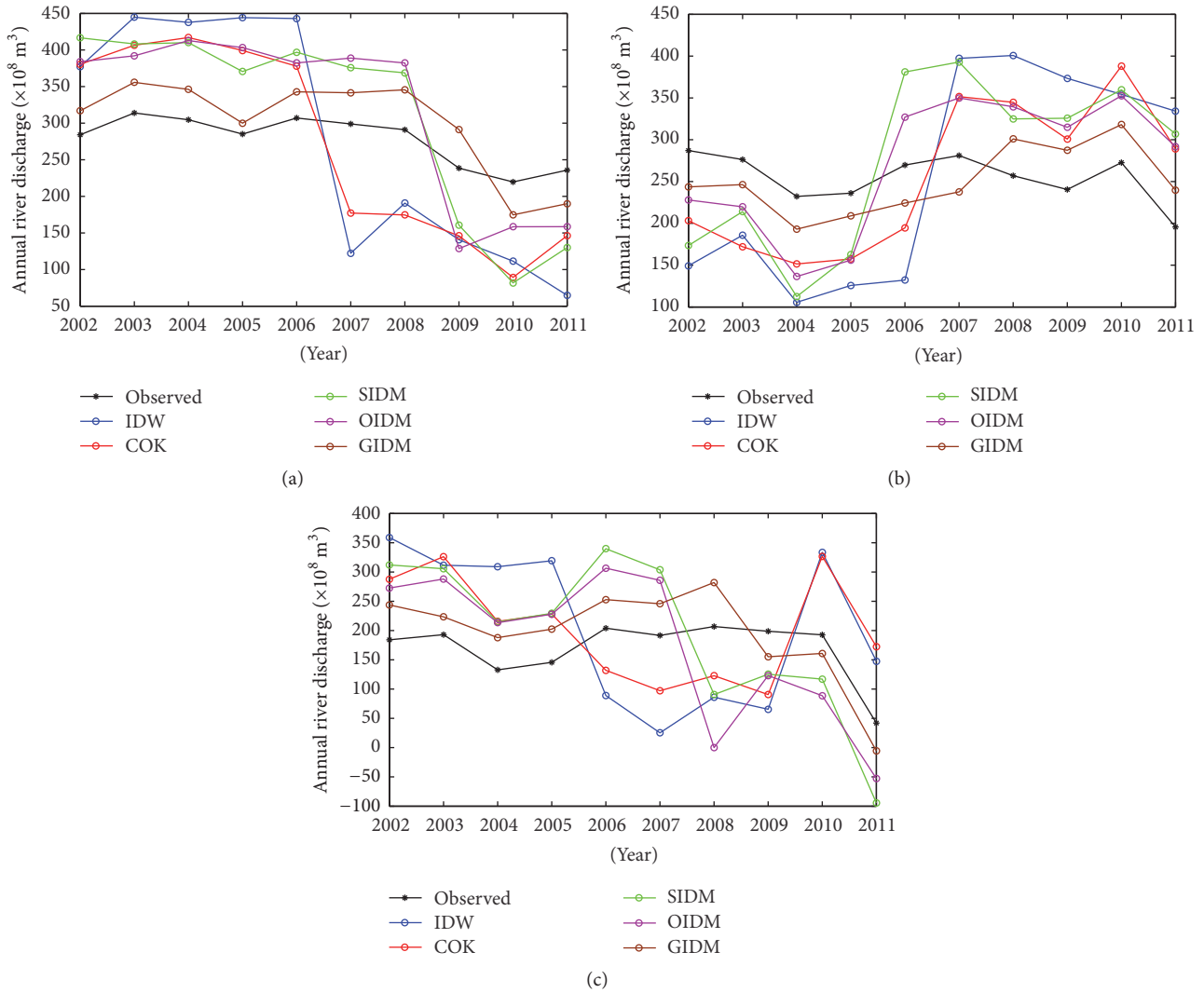
(a)

(b)

(c)

Figure 3: Comparison of observed runoff and interpolated runoff forced by IDW, COK, SIDM, OIDM, and GIDM at gauges Lanzhou (a), Huayuankou (b), and Lijin (c) from 2002 to 2011.

can be obtained through spatial interpolation. The actual annual runoff data of Hongqi and Baimasi site can be used to test the results of spatial interpolation. Through similar calculation steps as experiment 1, the annual natural runoff of two sites from 2002 to 2011 can be obtained.

Figure 4 describes reconstructed and observed runoff from different models. The correlation of the GIDM between reconstructed and observed runoff is best among the five different models, although some slight oversimulations and undersimulations exist. Table 5 shows a comparison among different five models and the GIDM method obtains best accuracy among five different models in terms of different evaluation measures. The averages of RMSE, R, E, and MAPE of the GIDM are 2.975 ($\times 10^8$ m^3), 0.9184, 0.8197, and 32.16%. Comparing with the average results of all models in Table 4, we can see that the interpolating results of all models in experiment 3 are worse than those in experiment 2, which also has only six training samples. For example, the average

MAPE of GIDM is 32.16% in experiment 3, while that in experiment 2 is only 20.50%. The reason is that the sites in experiment 3 on the tributaries of the Yellow River also belong to different subrivers, like the Hongqi site belonging to Tao subriver, Huangfu belonging to Huangfuchuan subriver, Wenjiachuan belonging to kuye subriver and Baimasi belonging to Beiluo subriver, and so forth. So the underlying surface and river hydrological processes are more complex, which will make the interpolation and prediction more difficult. Also, considering the modelling data, the data magnitudes of nine sites on the mainstream of the Yellow River in experiments 1 and 2 are not greatly different from each other, while the data magnitudes of eight sites on the tributaries of the Yellow River in experiment 3 are very different from each other. For example, the maximum ten-year average annual runoff is Huaxian, which is 52.89 ($\times 10^8$ m^3), and the minimum ten-year average annual runoff is Huangfuchuan, which is only 0.3022 ($\times 10^8$ m^3). The large differences among

TABLE 5: The RMSE, R, E, and MAPE of five different models in Hongqi site and Baimasi site (where boldface font indicates the best performance).

| Site	Model	RMSE	R	E	MAPE
Hongqi	IDW	6.42	0.8691	0.7572	92.55%
	COK	5.32	0.8951	0.8103	67.32%
	SIDM	5.61	0.9037	0.8239	69.14%
	OIDM	5.01	0.9124	0.8311	53.68%
	GIDM	**3.91**	**0.9251**	**0.8317**	**38.02**%
Baimasi	IDW	3.24	0.8238	0.7143	93.13%
	COK	2.91	0.8451	0.7326	65.28%
	SIDM	3.03	0.8572	0.7469	69.10%
	OIDM	2.80	0.8913	0.7905	44.89%
	GIDM	**2.04**	**0.9116**	**0.8076**	**26.29**%
The average	IDW	4.83	0.8465	0.7358	92.84%
	COK	4.115	0.8701	0.7715	66.30%
	SIDM	4.32	0.8805	0.7854	69.12%
	OIDM	3.905	0.90185	0.8108	49.29%
	GIDM	**2.975**	**0.9184**	**0.8197**	**32.16**%
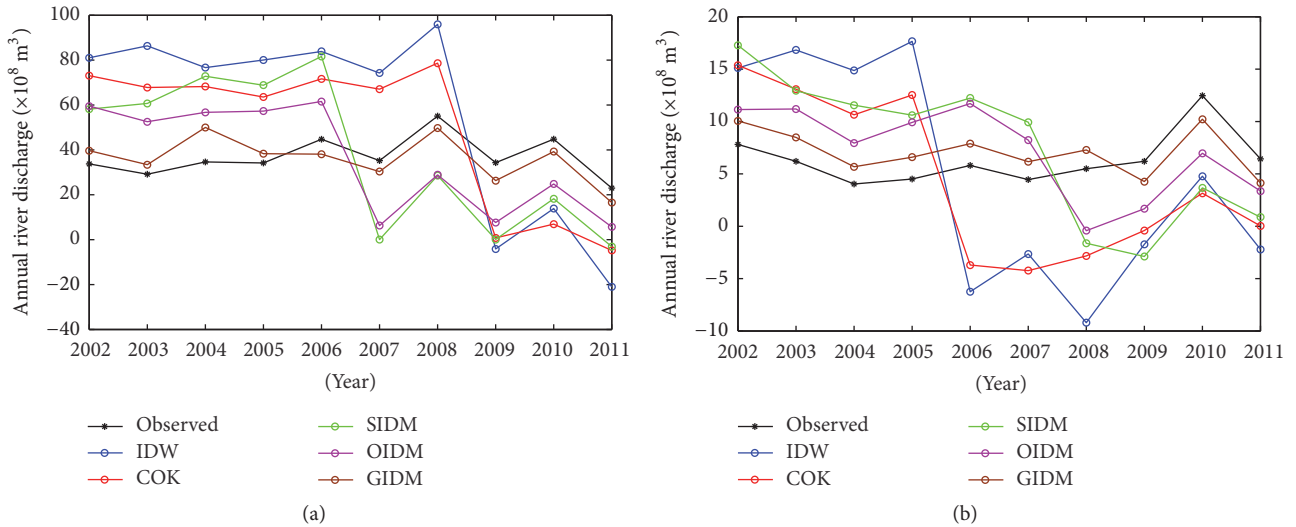
FIGURE 4: Comparison of observed runoff and interpolated runoff forced by IDW, COK, SIDM, OIDM, and GIDM at gauges Hongqi (a) and Baimasi (b) from 2002 to 2011.

the runoff magnitudes in different sites can also affect the interpolation results. But the average MAPE of GIDM is 32.16% and the average R is above 0.9 in experiment 3, which indicates that the general results of GIDM still could be acceptable.

4.3.2. Experiment 4: Interpolating Annual Runoff of Four Stations Using the Data of Other Four Stations. In experiment 4, from 2002 to 2011, the annual natural runoff data of four sites (Huangfu, Wenjiachuan, Baijiachuan, and Zhangjiashan) on the tributaries of the Yellow River are known for modelling, while the annual natural runoff data of four sites (Hongqi, Ganguyi, Baimasi, and Huaxian) on the tributaries of the Yellow River are seen as unknown for testing (Figure 1), which can be obtained through spatial interpolation. The actual

annual runoff data of Hongqi, Ganguyi, Baimasi, and Huaxian site can be used to test the results of spatial interpolation. Through similar calculation steps as experiment 1, the annual natural runoff of four sites from 2002 to 2011 can be obtained.

Figure 5 describes reconstructed and observed runoff from different models. The correlation of the GIDM between reconstructed and observed runoff is best among the five different models, although some slight oversimulations and undersimulations exist. Table 6 shows a comparison among different five models and the GIDM method obtains best accuracy among five different models in terms of different evaluation measures. The averages of RMSE, R, E, and MAPE of the GIDM are 3.21 ($\times 10^8$ m^3), 0.9033, 0.7891, and 39.32%. Comparing with the average results of all models in Table 5, we can see the interpolating results of all models

TABLE 6: The RMSE, R, E, and MAPE of five different models in Hongqi site, Ganguyi, Baimasi, and Huaxian site (where boldface font indicates the best performance).

Site	Model	RMSE	R	E	MAPE
Hongqi	IDW	7.88	0.8245	0.7123	116.43%
	COK	6.89	0.8678	0.7519	94.12%
	SIDM	6.97	0.8732	0.7644	97.28%
	OIDM	5.86	0.8905	0.7847	65.50%
	GIDM	**4.78**	**0.9033**	**0.7945**	**44.76%**
Ganguyi	IDW	2.45	0.5103	0.4016	156.52%
	COK	1.87	0.6722	0.5518	115.06%
	SIDM	1.94	0.6903	0.5802	120.67%
	OIDM	1.41	0.7388	0.6155	83.48%
	GIDM	**0.75**	**0.8792**	**0.7616**	**48.57%**
Baimasi	IDW	4.55	0.7891	0.6618	126.55%
	COK	3.67	0.8105	0.7025	98.43%
	SIDM	3.91	0.8145	0.7056	101.16%
	OIDM	3.34	0.8534	0.7501	72.05%
	GIDM	**2.56**	**0.884**	**0.7711**	**38.30%**
Huaxian	IDW	9.01	0.7365	0.6140	88.04%
	COK	7.55	0.7890	0.6658	61.28%
	SIDM	7.84	0.8104	0.7022	63.45%
	OIDM	6.06	0.8615	0.7507	45.18%
	GIDM	**4.75**	**0.9067**	**0.7890**	**25.65%**
The average	IDW	5.97	0.7151	0.5974	121.89%
	COK	4.99	0.7849	0.6680	92.22%
	SIDM	5.17	0.7971	0.6881	95.64%
	OIDM	4.17	0.8361	0.7253	66.55%
	GIDM	**3.21**	**0.8933**	**0.7791**	**39.32%**

are worse in experiment 4 than those in experiment 3. That is because in experiment 4 there are only 4 sites to train and the samples include less information of the runoff. But comparing experiment 3 and 4, we can see that the degree of the test results variation of the other four models is much larger than that of the GIDM method. The test results of the GIDM method only slightly deteriorate. For example, the MAPE of OIDM in Table 6 is 66.30%, which is a 34.51% increase than that of OIDM in Table 5, while the MAPE of GIDM in Table 6 is 39.32%, which is a 22.26% increase than that of GIDM in Table 5. Similarly, the reductions of R and E of GIDM are about 2.81% and 5.21%, which also indicate that the results of GIDM in experiment 4 are just a little worse than those in experiment 3.

The average MAPE of GIDM is 32.16% and the average R is above 0.9, which indicates that the general results of GIDM still could be acceptable. Considering only 4 sample sites for modelling, the spatial interpolation of the GIDM method is still good.

4.4. Experiments on Estimating Annual Runoff Time Series at Stations on the Mainstream and Tributaries of the Yellow River. In this section, two real examples of the annual runoff data ($\times 10^8$ m^3) from 2002 to 2011, taken at 17 stations on the mainstream and tributaries of the Yellow River (Figure 1) are shown to illustrate the implementation on the mainstream

and tributaries of the same watershed mixed together by different models.

4.4.1. Experiment 5: Interpolating Annual Runoff of Three Stations Using the Data of Other 14 Stations. In experiment 5, from 2002 to 2011, the annual natural runoff data of 14 sites (Tannaihai, Toudaoguai, Longmen, Huangfu, Wenjiachuan, Baijiachuan and Zhangjiashan, etc., 7 sites on the mainstream of the Yellow River and 7 sites on the tributaries of the Yellow River) are known for modelling, while the annual natural runoff data of three sites (Lanzhou, Lijin, and Hongqi, 2 sites on the mainstream of the Yellow River and 1 site on the tributaries of the Yellow River) are seen as unknown for testing (Figure 1), which can be obtained through spatial interpolation. The actual annual runoff data of Lanzhou, Lijin, and Hongqi sites can be used to test the results of spatial interpolation. Through similar calculation steps as experiment 1, the annual natural runoff of three sites from 2002 to 2011 can be obtained.

Figure 6 describes reconstructed and observed runoff from different models. The correlation of the GIDM between reconstructed and observed runoff is best among the five different models, although some slight oversimulations and undersimulations exist. Table 7 shows a comparison among different five models and the GIDM method obtains best accuracy among five different models in terms of different
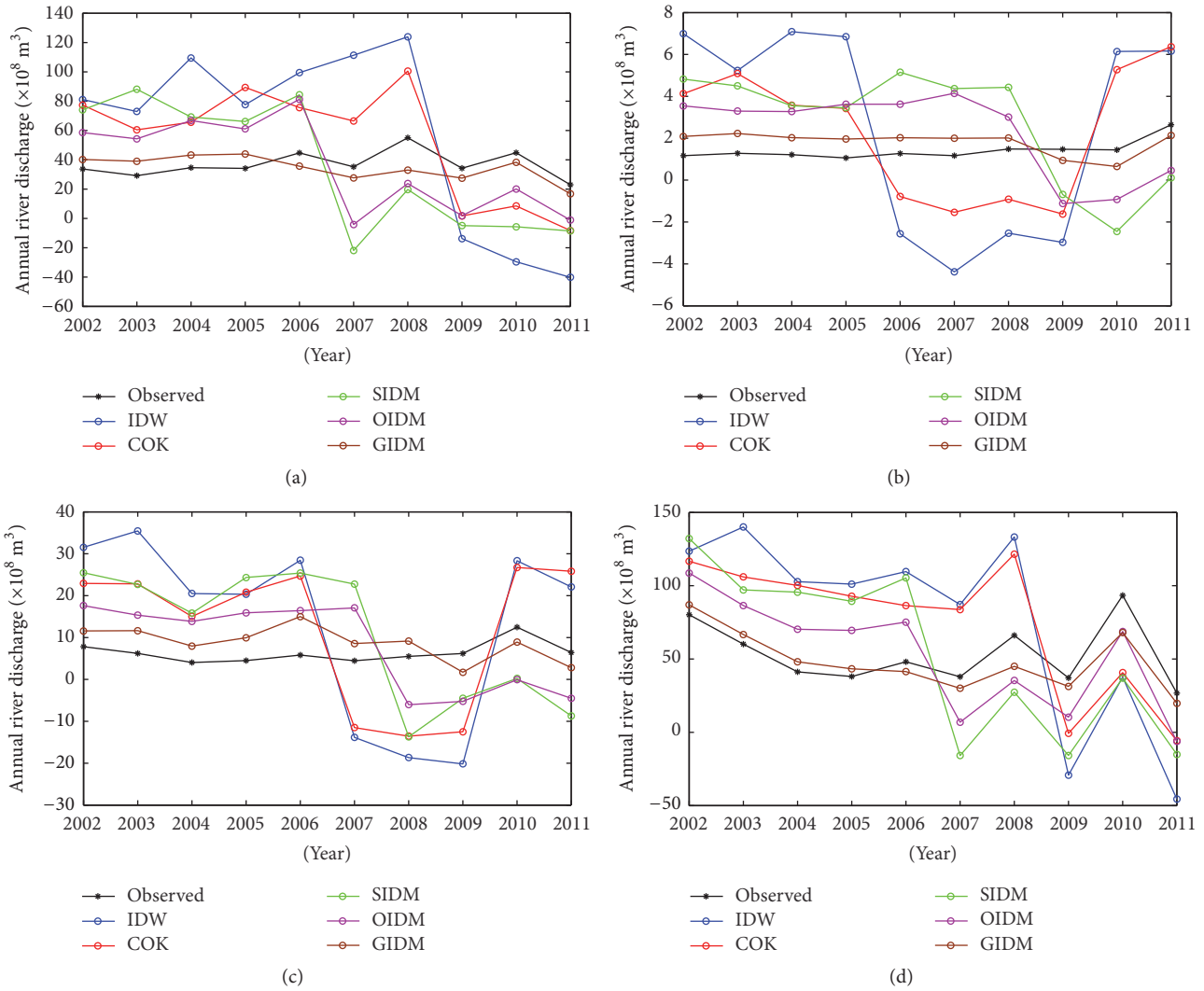
FIGURE 5: Comparison of observed runoff and interpolated runoff forced by IDW, COK, SIDM, OIDM, and GIDM at gauges Hongqi (a), Ganguyi (b), Baimasi (c) and Huaxian (d) from 2002 to 2011.

evaluation measures. The average of RMSE, R, E, and MAPE of the GIDM is 5.86 ($\times 10^8$ m^3), 0.9348, 0.8277, and 25.16%. Comparing with the average results of all models in Table 5, we can see the interpolating results of all models are better in experiment 5 than those in experiment 3. For example, the average MAPE of GIDM is 25.16% in experiment 5, while that in experiment 3 is only 32.16%. That is because similar as the experiment 3, the underlying surface and river hydrological processes of the mainstream and tributaries of the Yellow River mixed together are more complex, which will make the interpolation and prediction more difficult. And the large differences among the runoff in different sites can also affect the interpolation results. But the sample sites in experiment 5 are many, 14 sites, while the test sites are only three. More samples contain the more information, which will cause the final interpolation and prediction results to be better than those in experiment 3. The average MAPE of GIDM is 25.16% and the average R is above 0.9, which

indicates that the general results of GIDM are good and still could be acceptable.

4.4.2. Experiment 6: Interpolating Annual Runoff of Six Stations Using the Data of Other 11 Stations. In experiment 6, from 2002 to 2011, the annual natural runoff data of 11 sites (Tannaihai, Toudaoguai, Longmen, Huangfu, Wenjiachuan and Baijiachuan, etc., 6 sites on the mainstream of the Yellow River and 5 sites on the tributaries of the Yellow River) are known for modelling, while the annual natural runoff data of six sites (Lanzhou, Huayuankou, Linjin, Hongqi, Ganguyi, and Huaxian, 3 sites on the mainstream of the Yellow River and 3 sites on the tributaries of the Yellow River) are seen as unknown for testing (Figure 1), which can be obtained through spatial interpolation. The actual annual runoff data of Lanzhou, Huayuankou, Linjin, Hongqi, Ganguyi, and Huaxian site can be used to test the results of spatial interpolation. Through similar calculation steps as

TABLE 7: The RMSE, R, E, and MAPE of five different models in Lanzhou site, Lijin site, and Hongqi site (where boldface font indicates the best performance).

Site	Model	RMSE	R	E	MAPE
Lanzhou	IDW	13.12	0.8551	0.7417	42.89%
	COK	11.29	0.8704	0.7658	34.52%
	SIDM	11.16	0.8913	0.7928	34.11%
	OIDM	8.88	0.9146	0.8108	27.59%
	GIDM	**6.55**	**0.9344**	**0.8215**	**12.41**%
Lijin	IDW	11.76	0.7857	0.678	55.15%
	COK	9.51	0.846	0.7401	43.54%
	SIDM	9.33	0.8773	0.776	42.44%
	OIDM	8.78	0.9104	0.8008	37.69%
	GIDM	**6.67**	**0.9355**	**0.8298**	**21.55**%
Hongqi	IDW	7.44	0.8567	0.7428	105.18%
	COK	6.43	0.8807	0.7701	92.74%
	SIDM	6.77	0.8924	0.7815	93.45%
	OIDM	5.23	0.9158	0.8103	64.98%
	GIDM	**4.35**	**0.9344**	**0.8318**	**41.51**%
The average	IDW	10.77	0.8325	0.7208	67.74%
	COK	9.08	0.8657	0.7587	56.93%
	SIDM	9.09	0.8870	0.7834	56.67%
	OIDM	7.63	0.9136	0.8073	43.42%
	GIDM	**5.86**	**0.9348**	**0.8277**	**25.16**%

experiment 1, the annual natural runoff of six sites from 2002 to 2011 can be obtained.

Figure 7 describes reconstructed and observed runoff from different models. The correlation of the GIDM between reconstructed and observed runoff is best among the five different models, although some slight oversimulations and undersimulations exist. Table 8 shows a comparison among different five models and the GIDM method obtains best accuracy among five different models in terms of different evaluation measures. The averages of RMSE, R, E, and MAPE of the GIDM are 6.63 ($\times 10^8$ m^3), 0.8719, 0.7669, and 40.56%. Comparing with the average results of all models in Table 7, we can see the interpolating results of all models are worse in experiment 6 than those in experiment 5. That is because in experiment 6 the sample sites drastically reduce and the samples include less information of the runoff, which cause the interpolation results turning bad.

Also comparing with the average results of all models in Table 5, we can see the that interpolating results of all models are worse in experiment 6 than those in experiment 3. For example, the average MAPE of GIDM is 40.56% in experiment 6, while that in experiment 3 is only 32.16%. This shows that under the less training sample the interpolation on the mainstream and tributaries of the Yellow River mixed together will be more difficult than that in experiment 3. That is because the underlying surface and river hydrological processes of the mainstream and tributaries of the Yellow River mixed together are more complex than those in experiment 3, which will make the interpolation and prediction more difficult. Also, considering the modelling data, the differences of data magnitudes of 17

sites on the mainstream and tributaries of the Yellow River in experiments 6 (Difference of magnitude is 1000 times) are more than those of 8 sites on the tributaries of the Yellow River in experiment 3 (Difference of magnitude is 100 times). The large differences among the runoff magnitudes in different sites can also affect the interpolation results.

Previous studies also found that the spatial interpolation results on the mainstream and tributaries of the same river basin mixed together are slightly worse than those only on the mainstream or on the tributaries of the same river basin [24, 42, 43]. For example, Yan et al. [24] carried out a case study of the Huaihe river basin above Bengbu based on a hydrostochastic approach method. His study also found that the interpolation on the mainstream and tributaries of the Yellow River mixed together will be difficult. So in this experiment, the average MAPE of GIDM is 40.56% and the average R is above 0.85. Considering the less sample sites and the difficultly of the interpolation on the mainstream and tributaries of the Yellow River mixed together, the general results of GIDM are good and still could be acceptable.

4.5. Experiment 7: Experiments on Estimating Annual Runoff Time Series at Stations on Nonclosure Small Watershed of No Runoff Information. In this section, one real example of the annual runoff data ($\times 10^5$ m^3) from 2002 to 2011, taken at 5 stations in Daying mine region in Guizhou Province is shown to illustrate the implementation on the nonclosure small watershed of no runoff information by different models.

The Daying mine region is located in Dianzi village of Dafang town in Guizhou Province. Its geographic coordinates are [105°42′49″E-105°43′59″E, 27°12′57″-27°14′05″N]. The
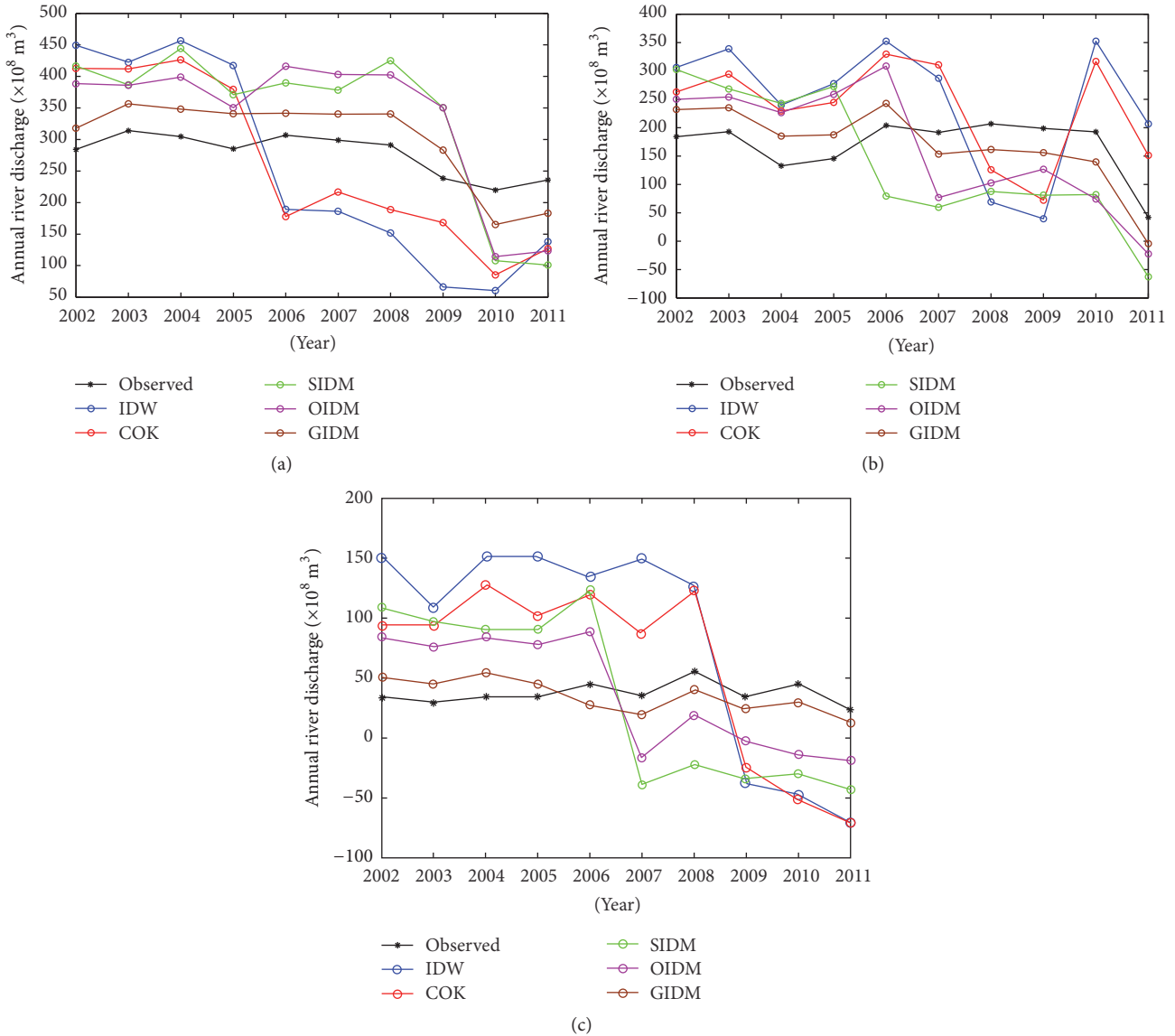
FIGURE 6: Comparison of observed runoff and interpolated runoff forced by IDW, COK, SIDM, OIDM, and GIDM at gauges Lanzhou (a), Lijin (b), and Hongqi (c) from 2002 to 2011.

Daying mine region is located in the western of Guizhou Plateau and belongs to north subtropical monsoon warm-wet weather patterns. The annual rainfall is about 1000 mm and the region belongs to rain source river basin. The Daying mine region belongs to the upstream basin of the Wuxi River and there is only Liulong hydrological station in this region. In this experiment, the Daying mine region can be seen as nonclosure small watershed of no runoff information (the Liulong sites can be seen as a test site). There are four hydrological sites (Yusha, Baina, Gaofeng, and Xiangshui) nearby the mine basin. The annual runoff data of each hydrological station have been reorganized and comprehensively checked, which makes the data available and reliable. Similarly as the experiment 1, from 2002 to 2011, the annual natural runoff data of 4 sites (Yusha, Baina,

Gaofeng, and Xiangshui) are known for modelling, while the annual natural runoff data of the Liulong site in the Daying mine region (nonclosure small watershed of no runoff information) are seen as unknown for testing, which can be obtained through spatial interpolation.

Table 9 shows a comparison among different five models and the GIDM method obtains best accuracy among five different models in terms of different evaluation measures. The averages of RMSE, R, E, and MAPE of the GIDM are 4.55 ($\times 10^5$ m^3), 0.9352, 0.8204, and 27.54%. Comparing with the average results of the above eight experiments, we can see for nonclosure small watershed of no runoff information, under the few training samples, the interpolating results of GIDM are still accurate, indicating that our model has good general applicability and reliability.
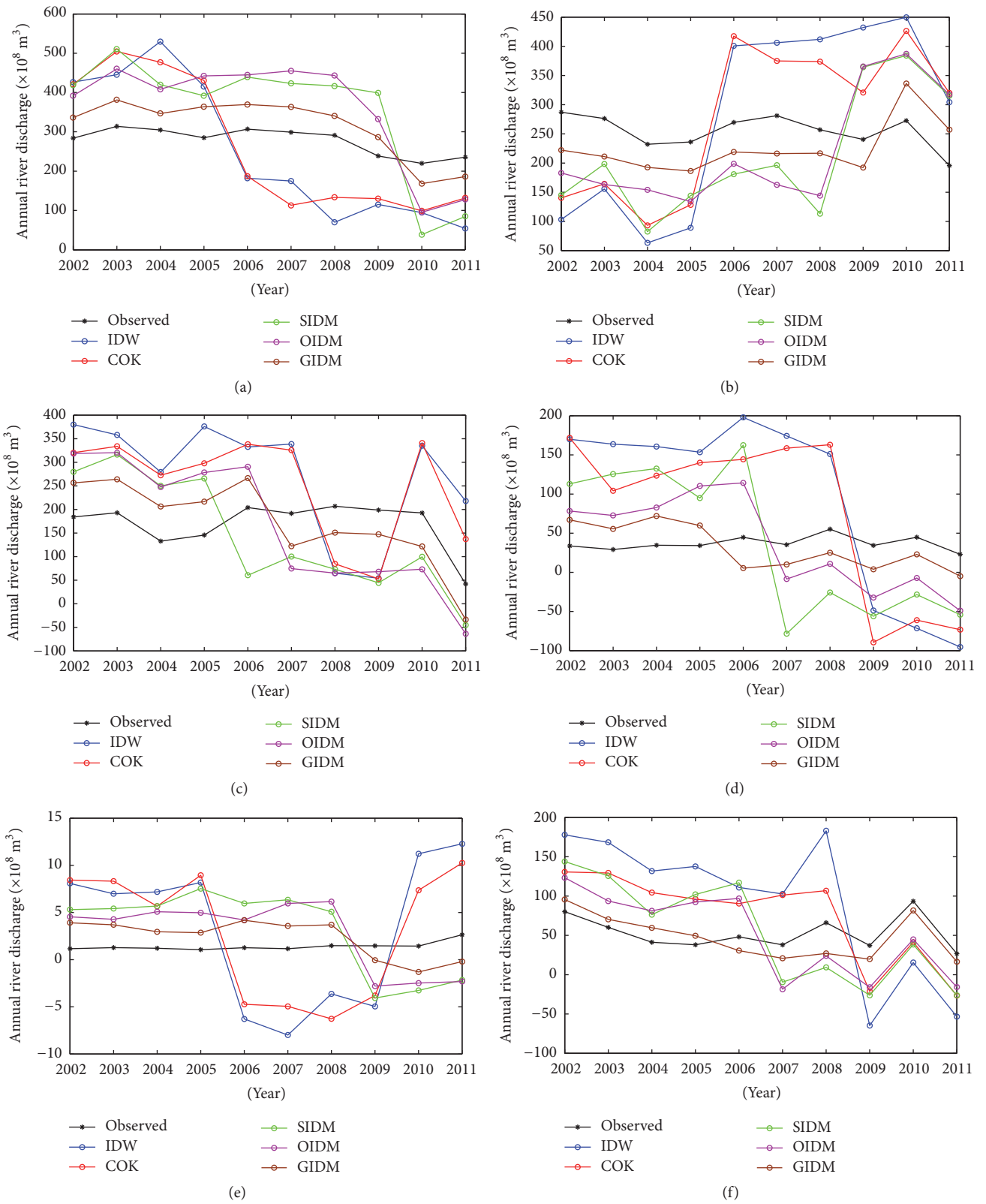
FIGURE 7: Comparison of observed runoff and interpolated runoff forced by IDW, COK, SIDM, OIDM, and GIDM at gauges Lanzhou (a), Huayuankou (b), Lijin (c), Hongqi (d), Ganguyi (e), and Huaxian (f) from 2002 to 2011.

Table 8: The RMSE, R, E, and MAPE of five different models in Lanzhou, Huayuankou, Lijin, Hongqi, Ganguyi, and Huaxian site (where boldface font indicates the best performance).

Site	Model	RMSE	R	E	MAPE
Lanzhou	IDW	17.33	0.8106	0.7051	52.12%
	COK	15.98	0.8333	0.7286	48.19%
	SIDM	14.17	0.8516	0.7455	45.15%
	OIDM	14.06	0.8914	0.7952	44.66%
	GIDM	**9.88**	**0.9004**	**0.8103**	**28.18%**
Huayuankou	IDW	17.88	0.8132	0.7045	49.15%
	COK	14.28	0.8518	0.7428	44.92%
	SIDM	14.44	0.8603	0.7519	45.85%
	OIDM	12.58	0.8884	0.7803	39.07%
	GIDM	**9.28**	**0.9036**	**0.8015**	**27.91%**
Lijin	IDW	15.71	0.7174	0.6140	65.41%
	COK	13.28	0.7521	0.6489	57.13%
	SIDM	13.89	0.8019	0.6977	58.62%
	OIDM	11.94	0.8144	0.7028	55.97%
	GIDM	**8.67**	**0.8607**	**0.7529**	**33.41%**
Hongqi	IDW	8.04	0.7917	0.6944	125.18%
	COK	7.02	0.8311	0.7207	96.31%
	SIDM	7.05	0.8357	0.7248	100.55%
	OIDM	6.21	0.8631	0.7550	77.42%
	GIDM	**5.14**	**0.8844**	**0.7706**	**54.28%**
Ganguyi	IDW	2.76	0.4782	0.3864	171.49%
	COK	2.09	0.6145	0.5028	124.57%
	SIDM	2.18	0.6455	0.5387	131.24%
	OIDM	1.67	0.6911	0.5907	96.72%
	GIDM	**1.02**	**0.8144**	**0.7058**	**66.18%**
Huaxian	IDW	9.85	0.7066	0.6024	94.27%
	COK	8.28	0.7422	0.6340	80.36%
	SIDM	8.33	0.7801	0.6705	83.27%
	OIDM	6.92	0.8288	0.7106	55.25%
	GIDM	**5.46**	**0.8677**	**0.7601**	**33.42%**
The average	IDW	11.93	0.7196	0.6178	92.94%
	COK	10.16	0.7708	0.6630	75.25%
	SIDM	10.01	0.7959	0.6882	77.45%
	OIDM	8.90	0.8295	0.7224	61.52%
	GIDM	**6.63**	**0.8719**	**0.7669**	**40.56%**

Table 9: The RMSE, R, E, and MAPE of five different models in Liulong site (where boldface font indicates the best performance).

Site	Model	RMSE	R	E	MAPE
Liulong	IDW	9.48	0.8440	0.7310	54.38%
	COK	8.86	0.8781	0.7640	46.67%
	SIDM	6.74	0.8804	0.7608	46.15%
	OIDM	5.61	0.9014	0.8011	35.45%
	GIDM	**4.55**	**0.9352**	**0.8204**	**27.54%**

5. Discussion and Conclusions

5.1. Discussion. Although Prediction in Ungauged Basins (PUB) in recent years has attracted more and more attention of research scientists and they made a lot of research, however, due to the problem complexity and methods limitations, the researches in this area still have much room to be improved. For example, Srinivasan et al. [44] propose a framework for developing Soil and Water Assessment Tool (SWAT) input data, including hydrography, terrain, land use, soil, tile, weather, and management practices, for the Upper Mississippi River basin (UMRB). The uncalibrated SWAT model ably predicts annual streamflow at 11 USGS gauges and crop yield at a four-digit hydrologic unit code (HUC) scale.

But there is an obvious error between their simulated results and the observed results. Recently, regionalization has widely used in the PUB (Jin et al., 2008); [45–47]. Regionalization is typically used to estimate parameter values of hydrological predictive tools for catchments without observed streamflow. But the prediction results are still not satisfactory and should be improved. For example, Jin et al. (2008) carried out regionalization study of a conceptual hydrological model in Dongjiang basin, south China. The hypothetical ungauged catchments produced acceptable results with an average coefficient of efficiency (ME) value equals to 0.72. Li et al. [47] propose a new regionalization method, called the index model, and predict flow duration curve in ungauged basins. The average coefficient of efficiency of their model is 0.722.

Comparing with the previous model, the advantages of our model are as follows:

(1) The GIDM model is a nonlinear information diffusion model, which is in line with the changes in the hydrological runoff.

(2) The demand of the training sites number is not high, such as the experiment 4. For a few training sample, the spatial structure simulation of hydrological variables is good and the interpolation results are accurate.

(3) The interpolation process of many geostatistical methods needs to add water balance constraints, which controls the export flows of each subbasin. Before the runoff calculation, the runoff data in the study should be normalized [22]. In their interpolation process, in order to measure the correlation between the basins, the distance between the subbasins of the interpolation algorithm should needs to be redefined [23]. The calculation of the entire process is complex and the conversion is troublesome. But the calculation process of our GIDM model is simple and convenient to operate.

(4) The test shows that the interpolation results of our model for the watershed intensely influenced by human activities (such as the Lijin sites in experiments 1 and 2) are also very good.

5.2. Conclusions. Our paper proposes a new method for reconstructing the river runoff based on incomplete data. The paper provides a new method to solve the Prediction in Ungauged Basins (PUB) problem. The algorithm can improve the IDMs by unraveling more information. Conventional IDMs, IDW, and COK are used for the comparison. Seven experiments based on GIDM for the runoff interpolation at different stations on the Yellow River are carried out, compared with SIDM, OIDM, IDW, and COK. These experiments can divide into five groups. The advantages of our method are as follows:

(1) GIDM is a useful tool for the river runoff spatial interpolation. The traditional method may get the acceptable result based on much more data. Through 6 experiments of the three species, we can fully verify that the interpolation results of the GIDM method not only are better than those of the traditional IDW and COK methods, but also can improve the traditional IDM interpolation by about 10–40%.

(2) In the middle and lower reaches of the Yellow River intensely influenced by human activities, the site observation data implies the impact of human activities on the natural runoff. Under this situation, the interpolation results of the GIDM method in study area are still good, which can overcome the traditional hydrological interpolation methods having good interpolating results only in the natural watershed.

(3) The previous six experiments proved that the interpolation results of the GIDM method on great rivers are good. Through experiment 7, we can see for nonclosure small watershed of no runoff information, under the few training samples, that the interpolating results of GIDM are still accurate, indicating that our model has good general applicability and reliability.

(4) Through the four different types of experiments, we can conclude that the number of modelling sites, the differences of the underlying surface and river hydrological processes among sites, and the difference of data magnitudes among sites are three key factors impacting the interpolation results of our GIDM method. This is why the interpolation and prediction results in experiments 6 are the worst among all the experiments. Even so, the average MAPE of GIDM in experiment 6 is under 45% and the average R is above 0.85, still in the acceptable range.

Although the new IDM is a useful tool for the river runoff spatial interpolation, it still cannot be acceptable with too small samples. So our future work will focus on these items. For example, we can establish a more useful diffusion function, or we can save computation time of the IDMs, and so forth.

Appendix

Principles of Information Diffusion and Simple Window Width

Let $X = \{x_1, x_2, \ldots, x_n\}$ be a given sample and let the universe of discourse be U. If and only if X is incomplete, there must be a reasonable information diffusion function $\mu(x_i, u)$, $u \in U$, which can accurately estimate the real relation $\mathbb{R}^1$. This is called the principle of information diffusion [25]. Let x_i ($i = 1, 2, \ldots, n$) be n independent identically distributed observations drawn from a population with density $p(x)$, $x \in \mathbb{R}^1$. Suppose $\mu(x)$ is a Borel measurable function in $(-\infty, +\infty)$.

$$\widehat{f}(x) = \frac{1}{nd} \sum_{i=1}^{n} \mu\left[\frac{x - x_i}{d}\right]. \tag{A.1}$$

The above equation is called an information diffusion estimator of $p(x)$ [25]. Huang [25] got the normal diffusion function as

$$\mu(x) = \frac{1}{d\sqrt{2\pi}} \exp\left[-\frac{x^2}{2d^2}\right] \qquad (A.2)$$

and the Simple Window Width (SWW) d as ($b = \max(X)$, $a = \min(X)$).

$$d = \begin{cases} 0.8146\,(b-a), & n = 5 \\ 0.5690\,(b-a), & n = 6 \\ 0.4560\,(b-a), & n = 7 \\ 0.3860\,(b-a), & n = 8 \\ 0.3362\,(b-a), & n = 9 \\ 0.2986\,(b-a), & n = 10 \\ \dfrac{2.6851\,(b-a)}{(n-1)}, & n \geq 11. \end{cases} \qquad (A.3)$$

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Acknowledgments

References

[1] M. Sivapalan, K. Takeuchi, S. W. Franks et al., "IAHS Decade on Predictions in Ungauged Basins (PUB), 2003-2012: shaping an exciting future for the hydrological sciences," *Hydrological Sciences Journal*, vol. 48, no. 6, pp. 857–880, 2003.

[2] G. Blöschl and M. Sivapalan, "Scale issues in hydrological modelling: a review," *Hydrological Processes*, vol. 9, no. 3-4, pp. 251–290, 1995.

[3] H. C. Riggs, "Low-flow investigations," in *Techniques of Water Resources Investigation of the U.S. Geological Survey*, vol. 4, U.S. Geological Survey, Washington, DC, USA, 1972.

[4] M. P. Mosley, "Delimitation of New Zealand hydrologic regions," *Journal of Hydrology*, vol. 49, no. 1-2, pp. 173–192, 1981.

[5] V. Klemeš, "Operational testing of hydrological simulation models," *Hydrological Sciences Journal*, vol. 31, no. 1, pp. 13–24, 1986.

[6] C.-Y. Xu, "Operational testing of a water balance model for predicting climate change impacts," *Agricultural and Forest Meteorology*, vol. 98-99, pp. 295–304, 1999.

[7] S. Guo, J. Wang, and J. Yang, "A semi-distributed hydrological model and its application in a macroscale basin in China," in *Soil-VegetationAtmosphere Transfer Schemes and Large-Scale Hydrological Models*, A. J. Dolman, A. J. Hall, M. L. Kavvas, T. Oki, and J. W. Pomeroy, Eds., pp. 167–174, IAHS, 2001.

[8] G. L. Vandewiele and A. Elias, "Monthly water balance of ungauged catchments obtained by geographical regionalization," *Journal of Hydrology*, vol. 170, no. 1-4, pp. 277–291, 1995.

[9] D. H. Burn and D. B. Boorman, "Estimation of hydrological parameters at ungauged catchments," *Journal of Hydrology*, vol. 143, no. 3-4, pp. 429–454, 1993.

[10] M. Huang, X. Liang, and Y. Liang, "A transferability study of model parameters for the variable infiltration capacity land surface scheme," *Journal of Geophysical Research D: Atmospheres*, vol. 108, no. 22, 2003.

[11] F. A. Abdulla and D. P. Lettenmaier, "Development of regional parameter estimation equations for a macroscale hydrologic model," *Journal of Hydrology*, vol. 197, no. 1–4, pp. 230–257, 1997.

[12] C.-Y. Xu, "Estimation of parameters of a conceptual water balance model for ungauged catchments," *Water Resources Management*, vol. 13, no. 5, pp. 353–368, 1999.

[13] W. Fernandez, R. M. Vogel, and A. Sankarasubramanian, "Regional calibration of a watershed model," *Hydrological Sciences Journal*, vol. 45, no. 5, pp. 689–707, 2000.

[14] E. H. Isaaks and R. M. Srivastava, *An Introduction to Applied Geostatistics*, Oxford University Press, New York, NY, USA, 1989.

[15] P. Goovaerts, *Geostatistics for Natural Resources Evaluation*, Oxford University Press, New York, NY, USA, 1997.

[16] G. De Marsily and S. Ahmed, "Application of kriging techniques in groundwater hydrology," *Journal of the Geological Society of India*, vol. 29, no. 1, pp. 57–82, 1987.

[17] J. O. Skøien, R. Merz, and G. Blöschl, "Top-kriging—geostatistics on stream networks," *Hydrology and Earth System Sciences*, vol. 10, no. 2, pp. 277–287, 2006.

[18] J. O. Skøien and G. Blöschl, "Spatiotemporal topological kriging of runoff time series," *Water Resources Research*, vol. 43, no. 9, Article ID W09419, 2007.

[19] G. Laaha, J. O. Skøien, and G. Blöschl, "Spatial prediction on river networks: comparison of top-kriging with regional regression," *Hydrological Processes*, vol. 28, no. 2, pp. 315–324, 2014.

[20] S. Castiglioni, A. Castellarin, A. Montanari, J. O. Skøien, G. Laaha, and G. Blöschl, "Smooth regional estimation of low-flow indices: physiographical space based interpolation and top-kriging," *Hydrology and Earth System Sciences*, vol. 15, no. 3, pp. 715–727, 2011.

[21] E. Sauquet, L. Gottschalk, and E. Leblois, "Mapping average annual runoff: a hierarchical approach applying a stochastic interpolation scheme," *Hydrological Sciences Journal*, vol. 45, no. 6, pp. 799–815, 2000.

[22] A. Cazenave and J. Chen, "Time-variable gravity from space and present-day mass redistribution in the Earth system," *Earth and Planetary Science Letters*, vol. 298, no. 3-4, pp. 263–274, 2010.

[23] M. J. M. Cheema and W. G. M. Bastiaanssen, "Local calibration of remotely sensed rainfall from the TRMM satellite for different periods and spatial scales in the Indus Basin," *International Journal of Remote Sensing*, vol. 33, no. 8, pp. 2603–2627, 2012.

[24] Z. Yan, J. Xia, and L. Gottschalk, "Mapping runoff based on water balance: a case study of the Huaihe River Basin above Bengbu," *Acta Geographica Sinica*, vol. 65, no. 7, pp. 841–852, 2010.

[25] C. Huang, "Principle of information diffusion," *Fuzzy Sets and Systems*, vol. 91, no. 1, pp. 69–90, 1997.

[26] C. Huang, "Information matrix and application," *International Journal of General Systems. Methodology, Applications, Education*, vol. 30, no. 6, pp. 603–622, 2001.

[27] Q. Li, J. Zhou, D. Liu, and X. Jiang, "Research on flood risk analysis and evaluation method based on variable fuzzy sets and information diffusion," *Safety Science*, vol. 50, no. 5, pp. 1275–1283, 2012.

[28] C. Huang, "Information diffusion techniques and small-sample problem," *International Journal of Information Technology & Decision Making*, vol. 1, no. 2, pp. 229–249, 2002.

[29] L. H. Feng and C. F. Huang, "A risk assessment model of water shortage based on information diffusion technology and its application in analyzing carrying capacity of water resources," *Water Resources Management*, vol. 22, no. 5, pp. 621–633, 2008.

[30] X. Z. Wang, Y. Yangsheng, and T. Yongjing, "The theory of optimal information diffusion estimation and its application," *Geospatial Information*, vol. 1, pp. 10–17, 2003.

[31] D. E. Goldberg and J. H. Holland, "Genetic algorithms and machine learning," *Machine Learning*, vol. 2, pp. 95–99, 1988.

[32] C. Bai, M. Hong, D. Wang, R. Zhang, and L. Qian, "Evolving an information diffusion model using a genetic algorithm for monthly river discharge time series interpolation and forecasting," *Journal of Hydrometeorology*, vol. 15, no. 6, pp. 2236–2249, 2014.

[33] R. Palm, "Multiple-step-ahead prediction in control systems with Gaussian process models and TS-fuzzy models," *Engineering Applications of Artificial Intelligence*, vol. 20, no. 8, pp. 1023–1035, 2007.

[34] W. Pei-Zhuang, "A factor spaces approach to knowledge representation," *Fuzzy Sets and Systems*, vol. 36, no. 1, pp. 113–124, 1990.

[35] R. S. V. Teegavarapu, M. Tufail, and L. Ormsbee, "Optimal functional forms for estimation of missing precipitation data," *Journal of Hydrology*, vol. 374, no. 1-2, pp. 106–115, 2009.

[36] A. R. Solow and S. M. Gorelick, "Estimating monthly streamflow values by cokriging," *Mathematical Geology*, vol. 18, no. 8, pp. 785–809, 1986.

[37] R. J. Hoeksema, R. B. Clapp, A. L. Thomas, A. E. Hunley, N. D. Farrow, and K. C. Dearstone, "Cokriging model for estimation of water table elevation," *Water Resources Research*, vol. 25, no. 3, pp. 429–438, 1989.

[38] T. S. Hu, K. C. Lam, and S. T. Ng, "River flow time series prediction with a range-dependent neural network," *Hydrological Sciences Journal*, vol. 46, no. 5, pp. 729–745, 2001.

[39] W.-C. Wang, K.-W. Chau, C.-T. Cheng, and L. Qiu, "A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series," *Journal of Hydrology*, vol. 374, no. 3-4, pp. 294–306, 2009.

[40] P. C. Nayak, K. P. Sudheer, D. M. Rangan, and K. S. Ramasastri, "A neuro-fuzzy computing technique for modeling hydrological time series," *Journal of Hydrology*, vol. 291, no. 1-2, pp. 52–66, 2004.

[41] J. E. Nash and J. V. Sutcliffe, "River flow forecasting through conceptual models part I—a discussion of principles," *Journal of Hydrology*, vol. 10, no. 3, pp. 282–290, 1970.

[42] W. Hao, W. Chengming, W. Jianhua et al., "Theory of runoff spatial distribution and the application in Wudinghe Basin," *Science in China: Series E*, vol. 34, supplement 1, pp. 74–86, 2004.

[43] L. Gottschalk and I. Krasovskaia, "Interpolation of annual runoff to grid networks applying objective methods//Macroscale Modelling of the Hydrosphere," *IAHS Publications*, vol. 214, pp. 81–90, 1993.

[44] R. Srinivasan, X. Zhang, and J. Arnold, "SWAT ungauged: hydrological budget and crop yield predictions in the upper Mississippi River basin," *Transactions of the ASABE*, vol. 53, no. 5, pp. 1533–1546, 2010.

[45] M. C. Peel, F. H. S. Chiew, A. W. Western, and T. A. McMahon, "Extension of Unimpaired Monthly Streamflow Data and Regionalisation of Parameter Values to Estimate Streamflow in Ungauged Catchments," National Land and Water Resources Audit, 2000, http://people.eng.unimelb.edu.au/mpeel/NLWRA.pdf.

[46] V. Y. Smakhtin, D. A. Hughes, and E. Creuse-Naudin, "Regionalization of daily flow characteristics in part of the Eastern Cape, South Africa," *Hydrological Sciences Journal*, vol. 42, no. 6, pp. 919–936, 1997.

[47] M. Li, Q. Shao, L. Zhang, and F. H. S. Chiew, "A new regionalization approach and its application to predict flow duration curve in ungauged basins," *Journal of Hydrology*, vol. 389, no. 1-2, pp. 137–145, 2010.