

# Streamflow prediction in ungauged basins by regressive regionalization: a case study in Huai River Basin, China

Jiyun Song, Jun Xia, Liping Zhang, Zhi-Hua Wang, Hui Wan and Dunxian She

## ABSTRACT

Streamflow information is of great significance for flood control, water resources utilization and management, ecological services, etc. Continuous streamflow prediction in ungauged basins remains a challenge, mainly due to data paucity and environmental changes. This study focuses on the modification of a nonlinear hydrological system approach known as the time variant gain model and the development of a regressive method based on the modified approach. This method directly correlates rainfall to runoff through physically based mathematical transformations without requiring additional information of evaporation or soil moisture. Also, it contains parsimonious parameters that can be derived from watershed properties. Both characteristics make this method suitable for practical uses in ungauged basins. The Huai River Basin of China was selected as the study area to test the regressive method. The results show that the proposed methodology provides an effective way to predict streamflow of ungauged basins with reasonable accuracy by incorporating regional watershed information (soil, land use, topography, etc.). This study provides a useful predictive tool for future water resources utilization and management for data-sparse areas or watersheds with environmental changes.

**Key words** | Huai River Basin, regressive regionalization, streamflow prediction, ungauged basins

**Jiyun Song**  
**Zhi-Hua Wang**  
School of Sustainable Engineering and the Built Environment,  
Arizona State University,  
Tempe, AZ 85287,  
USA

**Jiyun Song**  
**Jun Xia** (corresponding author)  
**Liping Zhang**  
**Hui Wan**  
**Dunxian She**  
State Key Laboratory of Water Resources and Hydropower Engineering Science,  
Wuhan University,  
Wuhan 430072,  
China  
E-mail: [xiajun666@whu.edu.cn](mailto:xiajun666@whu.edu.cn)

**Jun Xia**  
**Liping Zhang**  
**Hui Wan**  
**Dunxian She**  
Hubei Collaborative Innovation Center for Water Resources Security,  
Wuhan University,  
Wuhan 430072,  
China

## INTRODUCTION

Sustainable water resource management practices rely heavily on the accurate modeling of hydrological processes at watershed scales. In particular, numerical predictions of continuous streamflow is of paramount importance in a variety of fields, such as irrigation planning, flood control, engineering structure design, water resources utilization, and ecohydrological services (Parada & Liang 2010; Razavi & Coulibaly 2012; Cibir et al. 2014). In practical applications, however, we often need to deal with many ungauged or poorly gauged basins without adequate and accurate streamflow observations (Sivapalan et al. 2003). These data-sparse basins often exist in mountainous areas

(Castellari et al. 2007), unregulated regions (Stainton & Metcalfe 2007), and rural or remote areas (Makungo et al. 2010). Some gauged basins may also change to be ungauged when the previous streamflow information is no longer suitable to describe the hydrological responses to environmental changes, such as human-induced land use change. Furthermore, modified land surface processes in watersheds, e.g., heavily built terrains, will in turn influence the local hydroclimate via land-atmospheric interactions (Song & Wang 2015a, 2015b), thus resulting in higher uncertainty in predicting hydrological responses in the region.

In general, an effective hydrological prediction system consists of an appropriate model structure, a set of calibrated model parameters, and accurate model inputs. The hydrological system approach (Singh 1988) is found to be flexible as compared to other conceptual models in data-sparse areas under uncertainty perturbation and environmental change (Xia 1991; Xia *et al.* 2005). Xia (1991) developed a nonlinear hydrological system approach based on Volterra functional series, known as the time variant gain model (TVGM). The TVGM has been tested over ten different basins in China, Japan, the United States, and Australia (Xia *et al.* 1997), and found to be effective for daily and hourly streamflow forecasting under different climate conditions (semi-arid and humid) with parsimonious parameters. For better predictability, a modified TVGM with two runoff types is proposed here and selected as the hydrological modeling approach for ungauged catchments in this study.

Model parameters are often calibrated based on previous observation data, which are unavailable at ungauged or poorly gauged sites. The lack of model calibration and verification due to the paucity of measurement data therefore requires a different methodology for parameter estimation. Regionalization (Blöschl & Sivapalan 1995; Jin *et al.* 2009; Kizza *et al.* 2013) is a feasible way for predictions in ungauged basins by transferring hydrological information from gauged basins. There are three types of regionalization method: spatial proximity, physical similarity, and regression (Oudin *et al.* 2008). The spatial proximity method (Mosley 1981; Vandewiele & Elias 1995) focuses on the geographical similarity and employs the parameter values from the geographic neighbors without considering the heterogeneity of the catchments included. The physical similarity method (Reed 1999; Patil & Stieglitz 2012) is based on hydrological proximity, regardless of geographical location of the study area and the donor areas. The regression method (Abdulla & Lettenmaier 1997; Post & Jakeman 1999; Seibert 1999; Xu 1999, 2003; Li *et al.* 2010), in contrast, estimates the model parameters of ungauged catchments according to a posteriori relationships between catchment descriptors (both physical and climatic) and model parameter values calibrated at gauged sites. This approach is capable of incorporating more catchment information, but also exhibits more uncertainties in model parameter and catchment descriptor (Oudin *et al.* 2008).

Previous studies have attempted to compare the three regionalization approaches, without reaching a clear consensus. This is mainly because these studies were based on a variety of catchment sets, climatic situations, donor catchment sets, catchment descriptors, and hydrological models, and comparisons were not made on the same ground (Oudin *et al.* 2008). It is found that the best choice of regionalization approach is site specific rather than universal. With an appropriate hydrological model, regression approach works best in most warm temperate regions (Razavi & Coulibaly 2012). While our study area of the Huai River Basin, China is located in a warm temperate region, the choice of regression approach is appropriate.

Our objective in this study is to predict 3-hourly streamflow for ungauged catchments by regression approach based on the modified TVGM. The structure of this paper is arranged as follows. First, the study area and data collection are introduced. In the next section, the modified TVGM is described followed by its validated. Based on the modified TVGM, a parametric regression approach is established and applied for streamflow predictions. Then, results of the regression approach and discussions on its underlying physics are presented. Lastly, findings and implications of this article are presented.

---

## STUDY AREA AND AVAILABLE DATA

### Study area

Huai River Basin, as one of China's seven major river basins, is located between 30°55'–36°36' N and 111°55'–121°25'E, with a total drainage area of 270,000 km<sup>2</sup>. This basin belongs to a warm temperate semi-humid monsoon area and acts as the transition zone of climate between southern monsoon climate and northern continental climate in China (Chen *et al.* 2011). The annual average temperature is 11–16°C which increases gradually from north to south. The annual average water surface evaporation is 900–1,500 mm (Zhang *et al.* 2011). The long-term annual average precipitation is 911 mm with a decreasing trend from south to north, mountainous areas to plains, and coastal areas to inland. Watershed runoff is mainly

concentrated in the flood season and the long-term average annual runoff depth is 231 mm. The upstream of Bengbu hydrological station in Huai River Basin is a typical large-scale basin in China and is chosen as the study area in this paper, as shown in Figure 1.

### Available data

Considering the impacts of water projects in the study area, Zhang et al. (2010) showed that the flow in the non-flood season was reduced by 5%, while the change of flow in the flood season was not significant when the sluices along the river channels were kept open. To minimize the anthropogenic impacts, the continuous 3-hourly rainfall–runoff data from April to September (flood season) of year 2000–2008 period were collected at 13 hydrological stations in the study area: Huangchuan, Luohe, Jiangjiaji, Bantai, Mengcheng, Zhoukou, Jieshou, Xixian, Huaibin, Wangjiaba, Runheji, Lutaizi, and Bengbu (see Figure 1). The digital elevation model (DEM) datasets were obtained from the shuttle radar topography mission provided by the National Aeronautics and Space Administration and the National Imagery and Mapping Agency, USA, with a resolution of 90 m. The topography and spatial topology associated with large- and medium-scale basins are described in details by these datasets. The DEM data were used to delineate the catchments and analyze the topographic information of each catchment (see Table 1). The selected catchments include small, medium, and large basins with the drainage area ranging from 2,050 to 121,330 km<sup>2</sup>, as shown in Figure 2. We select the time period from 2000 to 2008 to represent different flow characteristics with high, median, and dry flow years. In addition, rainfall characteristics including long-term average rainfall volume and the location of rainfall centroid are also summarized in Table 1.

The soil-type dataset was obtained from the United Nations Food and Agriculture Organization (FAO) and the soil types were defined by global soil classification of FAO and UNESCO. Soil properties and soil water characteristics of different catchments are derived as presented in Tables 2 and 3, respectively. The land use type data with a scale of 1:1,000,000 were obtained from the Institute of Geographic

Sciences and Natural Resources Research, Chinese Academy of Sciences. The land use types in the study area can be divided into the following six types: forest (FRST), range of grassland and meadow (RNGE), water area (WATER), urban areas with high density (URHD), rice paddy (RICE), and agricultural dry land (AGRR). The ratios of different land use types in each catchment are presented in Table 4.

## METHODOLOGY

### TVGM

Taking into account the nonlinearity of a hydrological system, a rainfall–runoff process can be expressed by a second-order Volterra functional series (Xia 1991; Xia et al. 2005; Carassale & Kareem 2009) as:

$$y(t) = \int_0^L h(\tau)x(t-\tau)d\tau + \int_0^L \int_0^L g(\tau, \sigma)x(t-\tau)x(t-\sigma)d\tau d\sigma \quad (1)$$

where  $y$  is the system output (e.g., runoff),  $x$  is the system input (e.g., rainfall),  $h$  is a linear response function,  $g$  is a nonlinear response function,  $L$  is the system's memory length,  $t$ ,  $\tau$ ,  $\sigma$  are time variants. Equation (1) describes the nonlinear responses of a hydrological system, but is, in general, not analytically tractable.

Xia et al. (1997) proposed a relatively simplified nonlinear systematic rainfall–runoff model (i.e., the TVGM), equivalent to a special form of the complex second-order nonlinear Volterra model. The hydrological processes of the TVGM are given by:

$$R(t) = G(t)X(t) \quad (2)$$

$$Y(t) = \int_0^m U(\tau)R(t-\tau)d\tau \quad (3)$$

where  $R$  is the rainfall excess,  $G$  is a time variant gain coefficient related to soil moisture conditions,  $X$  is the hydrological system input (i.e., rainfall),  $U$  is the response function, and  $Y$  is the hydrological system output

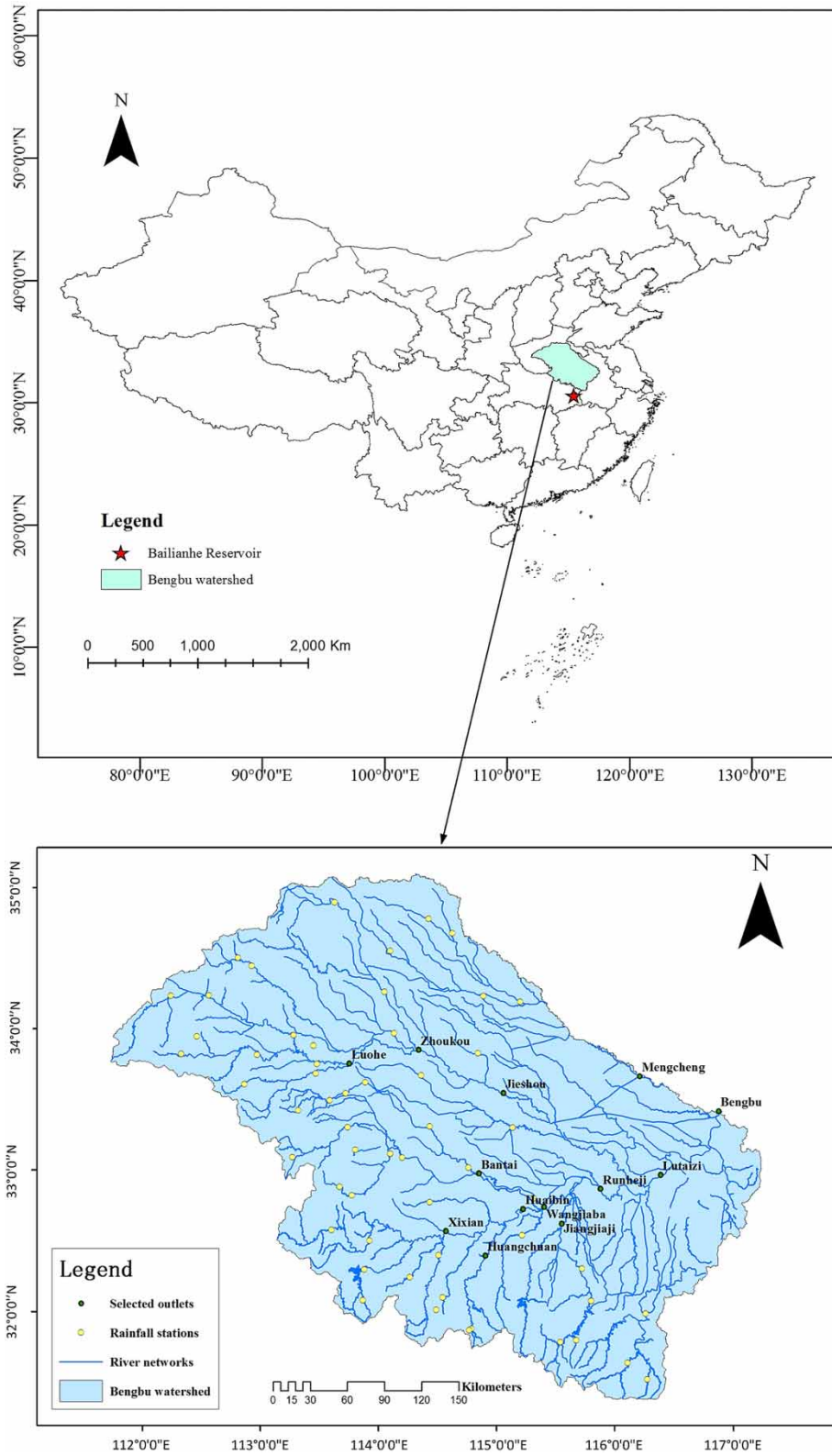
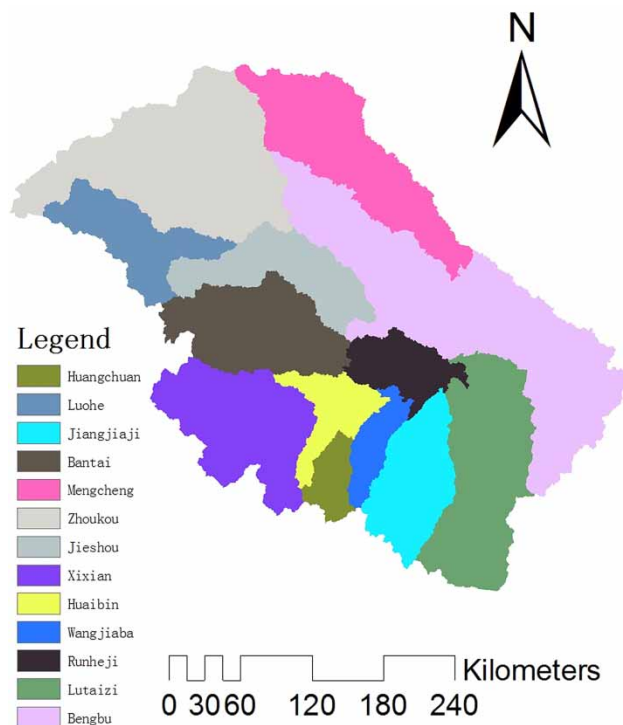


Figure 1 | The geographical location and hydrological information of the study area in China.

**Table 1** | Basic rainfall–runoff characteristics of each catchment

Catchment	A (km <sup>2</sup> )	N <sub>p</sub>	$\alpha$	top (m)	P (mm)	Mean rainfall centroid		
						LON	LAT	d (km)
Huangchuan	2,050	6	0.0045	19.02	931.61	115.05	32.13	45.68
Jiangjiaji	5,930	12	0.0052	20.42	803.94	115.73	32.3	56.39
Xixian	10,190	24	0.0029	22.2	848.28	114.73	32.33	72.21
Bantai	11,280	26	0.001	18.18	744.92	115.07	32.72	109.16
Luohe	12,150	12	0.0051	20.21	702.22	114.03	33.58	83.92
Mengcheng	15,475	13	0.0002	20.54	633.67	116.55	33.28	145.74
Huaibin	16,005	34	0.0022	22.67	798.58	115.42	32.43	111.5
Zhoukou	25,800	42	0.0031	21.31	453.79	114.65	33.63	151.11
Jieshou	29,290	45	0.0024	21.47	630.16	115.35	33.27	207.76
Wangjiaba	30,630	61	0.002	22.67	651.28	115.6	32.43	131.12
Runheji	40,360	68	0.0029	22.67	704.41	116.1	32.52	164.77
Lutaizi	88,630	91	0.0027	22.67	695.39	116.63	32.57	178.38
Bengbu	121,330	158	0.0018	23.18	720.05	117.38	32.93	272.93

A is the catchment area; N<sub>p</sub> is the number of dominant rainfall stations;  $\alpha$  is the mainstream slope; top is the topographical wetness index related to soil state; P is the long-term average precipitation; LON and LAT refer to the longitude and latitude of mean rainfall centroid respectively; d is the distance between the mean rainfall centroid and the corresponding hydrological station.

**Figure 2** | Distribution of 13 sub-watersheds in the study area (note that some large watersheds are overlaid by other smaller watersheds).**Table 2** | Mean soil characteristics of each catchment

Catchment	Soil layer depth (mm)	Sand (%)	Silt (%)	Clay (%)
Huangchuan	1,117.94	46.68	23.22	28.55
Jiangjiaji	766.05	46.83	26.03	21.36
Xixian	1,015.23	39.88	30.26	21.74
Bantai	1,162.59	38.08	33.69	22.97
Luohe	913.82	37.7	30.87	25.43
Mengcheng	968.36	44.42	26.44	22.68
Huaibin	1,027.46	37.59	30.97	24.15
Zhoukou	917.38	38.54	31.3	18.47
Jieshou	904.42	35.21	30.63	21.42
Wangjiaba	1,046.85	36.03	31.42	23.96
Runheji	964.23	35.38	29.63	23.19
Lutaizi	923.86	34.01	28.77	22.6
Bengbu	800.22	28.45	25.27	19.79

(i.e., runoff). Here  $G$  can be written as:

$$G(t) = g_1 + g_2 API(t) \quad (4)$$

where  $g_1$  and  $g_2$  are two parameters related to watershed properties that are not time variant,  $API$  is the time

**Table 3** | Mean soil water characteristics of each catchment

Catchment	Wilting point (% Vol.)	Field capacity (% Vol.)	Saturation (% Vol.)	Available water (mm/m)
Huangchuan	15.01	25.62	45.79	106.67
Jiangjiaji	13.26	23.82	44.08	105.83
Xixian	13.81	25.16	43.46	111.67
Bantai	15.15	28.18	44.03	130.00
Luohe	16.37	28.49	44.40	120.83
Mengcheng	14.04	23.75	45.25	96.67
Huaibin	14.83	26.59	43.86	116.67
Zhoukou	12.46	24.04	40.99	115.83
Jieshou	14.00	25.54	40.87	115.00
Wangjiaba	14.91	26.92	43.00	119.17
Runheji	14.43	25.86	41.54	114.17
Lutaizi	14.13	25.09	40.51	109.17
Bengbu	12.55	21.94	35.07	93.33

**Table 4** | Land use characteristics of each catchment

Catchment	FRST (%)	RNGE (%)	WATER (%)	URHD (%)	RICE (%)	AGRR (%)
Huangchuan	19	1	2	1	32	44
Jiangjiaji	32	12	3	1	35	18
Xixian	36	0	3	1	20	40
Bantai	9	2	3	1	1	84
Luohe	21	5	5	1	0	68
Mengcheng	1	1	3	0	1	95
Huaibin	26	0	3	1	24	46
Zhoukou	9	4	2	3	1	81
Jieshou	10	3	2	3	1	81
Wangjiaba	23	1	3	1	28	45
Runheji	19	3	3	1	20	54
Lutaizi	21	4	3	1	26	45
Bengbu	12	3	2	2	16	65

variant antecedent precipitation index, which can be simulated as a response of a simple linear reservoir to the rainfall  $X$  (Ahsan & O'Connor 1994; Xia et al. 1997, 2005) by:

$$API(t) = \int_0^t U_0(\sigma)X(t - \sigma)d\sigma \tag{5}$$

Here,  $U_0$  is a response function given by:

$$U_0(\sigma) = \frac{1}{K_e} \exp\left(-\frac{\sigma}{K_e}\right) \tag{6}$$

where  $K_e$  is a parameter indicating the rate of soil moisture recession.

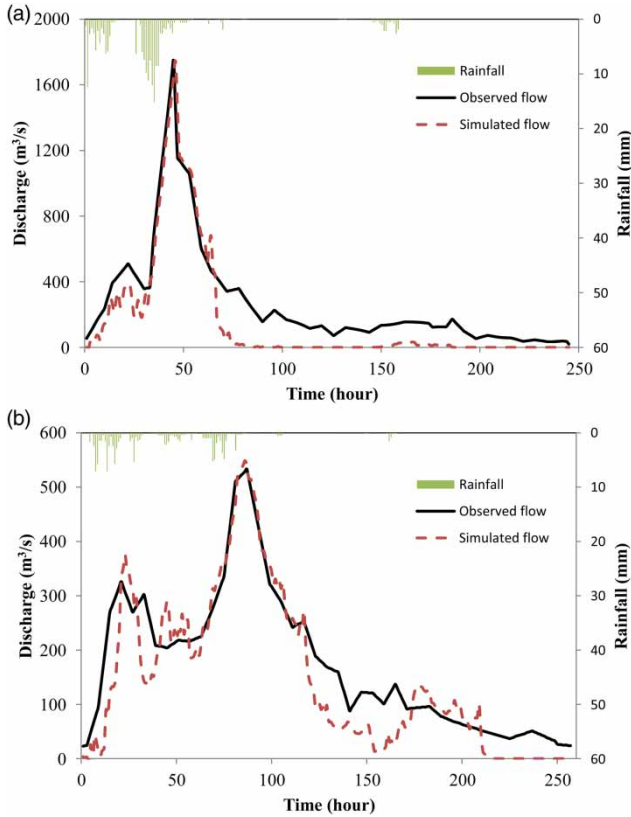
Substituting Equation (5) into Equation (4), and then combining with Equation (2), we have:

$$R(t) = g_1X(t) + \int_0^t g_2U_0(\sigma)X(t - \sigma)X(t)d\sigma \tag{7}$$

Substituting Equation (7) into Equation (3), we obtain an isomorphic representation of the second-order Volterra model as:

$$Y(t) = \int_0^m g_1U(t - \tau)X(\tau)d\tau + \int_0^m \int_0^\tau g_2U_0(t - \sigma)U(t - \tau)X(\sigma)X(\tau)d\tau d\sigma \tag{8}$$

As a simple nonlinear system approach, the TVGM can be used for hydrological forecasting with reasonable accuracy even if catchment evapotranspiration or soil moisture data are unavailable. The TVGM was evaluated at the outlet of Bailianhe Reservoir, which controls a relatively small area of 1,797 km<sup>2</sup> with negligible human interference and close to Huai River Basin with climatic similarity (marked by a star in Figure 1). The model was applied in this area to simulate a unimodal flood event on July 18, 1982 and a bimodal flood event on August 19, 1982. By comparing the results of simulation with observations as shown in Figure 3, there is a good agreement for the high discharge range, while larger discrepancy is found in low streamflow periods. The reason is that the TVGM does not generate runoff without rainfall, as  $Y$  vanishes if  $X$  is zero according to Equation (8). This is not physically reasonable because baseflow still exists even if there is no precipitation. As indicated in Napiórkowski (1992), negative outflows can also be obtained through this model. Therefore, a modification is necessary in order to improve the predictability of the TVGM in low streamflow periods.



**Figure 3** | Flood simulation by original TVGM in Baillianhe Reservoir for (a) a unimodal flood event on July 18, 1982 and (b) a bimodal flood event on August 19, 1982.

**Modified TVGM**

Based on the original TVGM, here we propose an advanced TVGM by introducing two separate time variant gain coefficients for two runoff types, i.e., quick flow and baseflow, respectively. For quick flow:

$$R_s(t) = G_s(t)X(t) \tag{9}$$

where  $R_s$  is the effective rainfall for generating quick flow,  $G_s$  is a time variant gain coefficient for quick flow and can be expressed as:

$$G_s(t) = g_1 + g_2API(t) \tag{10}$$

Similarly, baseflow is calculated by:

$$R_g(t) = G_g(t)API(t) \tag{11}$$

where  $R_g$  is the effective rainfall for generating baseflow,  $G_g$  is a time variant gain coefficient for baseflow and can be expressed as:

$$G_g(t) = g_3 + g_4API(t) \tag{12}$$

In Equations (10) and (12), coefficients  $g_1, g_2, g_3,$  and  $g_4$  are constants for a specific watershed.

In addition, unlike the flow routing process by using a single response function for all effective rainfall in the original TVGM as indicated in Equation (3), here we use two separate response functions for routing of quick flow  $Y_s$  and baseflow  $Y_g$ , respectively, in correspondence with two time variant gain flow generation mechanisms:

$$Y_s(t) = \int_0^m u_s(\tau)R_s(t - \tau)d\tau \tag{13}$$

$$Y_g(t) = \int_0^m u_g(\tau)R_g(t - \tau)d\tau \tag{14}$$

where  $u_s$  and  $u_g$  are the response functions for quick flow and baseflow, respectively. These functions are given by:

$$u_s(t) = \frac{1}{K\Gamma(n)} \left(\frac{t}{K}\right)^{n-1} e^{-t/K} \tag{15}$$

$$u_g(t) = \frac{1}{K_g} e^{-t/K_g} \tag{16}$$

Here,  $n$  is a numerical parameter indicating the capacity of watershed storage, which is equivalent to the number of linear reservoirs;  $K$  is a storage-discharge parameter with the dimension of time (Nash 1957; Young & Beven 1991); and  $K_g$  is the storage coefficient of ground water with the dimension of time.

Since rainfall data are usually recorded in discrete forms, the analytical forms of instantaneous response functions, i.e.,  $u_s, u_g,$  need to be practically discretized for a given duration. The quick flow response function  $u_s,$  viz. the instantaneous unit hydrograph can be discretized using S-curve method (Sherman 1932; Cleveland et al. 2008), while the baseflow response  $u_g$  can be converted into a discrete form through the water balance equation by assuming

a linear relationship between ground water storage and discharge (Pedersen *et al.* 1980; Purcell 2006). With commonly adopted unit hydrograph, S-curve and underground linear reservoir, the quick flow  $Y_s$  and baseflow  $Y_g$  can be calculated by discrete convolution integrals as:

$$Y_s(t) = c \sum_{i=1}^{T_u} u_s(\Delta t, i) R_s(t - i + 1) \quad (17)$$

$$Y_g(t) = cR_g(t - 1)(1 - KKG) + Y_g(t - 1)KKG \quad (18)$$

where  $t$  is the current time step,  $T_u$  is the duration of the  $u_s$ ,  $c = A/(3.6 \times \Delta t)$  is a coefficient to transform the unit of runoff volume from  $\text{m}^3 \text{s}^{-1}$  to mm,  $A$  is the catchment area in  $\text{km}^2$ ,  $\Delta t$  is the time interval in hour,  $R_s$  and  $R_g$  are effective rainfall in mm,  $KKG$  is a coefficient given by  $KKG = (K_g - 0.5\Delta t)/(K_g + 0.5\Delta t)$ .

### Derivation of model parameters

During the model calibration process, the quick flow  $Y_s$  and baseflow  $Y_g$  are first separated from observed streamflow by digital filter (Chapman 1999; Eckhardt 2005). Since a runoff process can be regarded as the distribution of effective rainfall (Beven 2011), there exists the following water balance equation in a certain duration of  $T_m$  for both quick flow and baseflow as:

$$\int_{t-T_m}^t Y_s(s) ds = \int_{t-T_m}^t R_s(s) ds \quad (19)$$

$$\int_{t-T_m}^t Y_g(s) ds = \int_{t-T_m}^t R_g(s) ds \quad (20)$$

where  $Y_s$  and  $Y_g$  are quick flow and baseflow, respectively.

Substituting Equation (9) into Equation (19), we have:

$$\begin{aligned} \int_{t-T_m}^t Y_s(s) ds &= \int_{t-T_m}^t G_s(s) X(s) ds = \int_{t-T_m}^t (g_1 + g_2 API(s)) X(s) ds \\ &= g_1 \int_{t-T_m}^t X(s) ds + g_2 \int_{t-T_m}^t API(s) X(s) ds \end{aligned} \quad (21)$$

Similarly, substituting Equation (11) into Equation (20), we have:

$$\begin{aligned} &\int_{t-T_m}^t Y_g(s) ds \\ &= \int_{t-T_m}^t G_g(s) API(s) ds = \int_{t-T_m}^t (g_3 + g_4 API(s)) API(s) ds \\ &= g_3 \int_{t-T_m}^t API(s) ds + g_4 \int_{t-T_m}^t API(s)^2 ds \end{aligned} \quad (22)$$

The four runoff generation parameters ( $g_1, g_2, g_3, g_4$ ) can be calculated via the Householder least square method (Businger & Golub 1965) according to Equations (21) and (22). The three parameters in flow routing process ( $n, K, KKG$ ) are calibrated by genetic algorithm (Rabunal *et al.* 2007; Chen *et al.* 2009) with the objective of least root mean square error between simulated and observed flow in this paper.

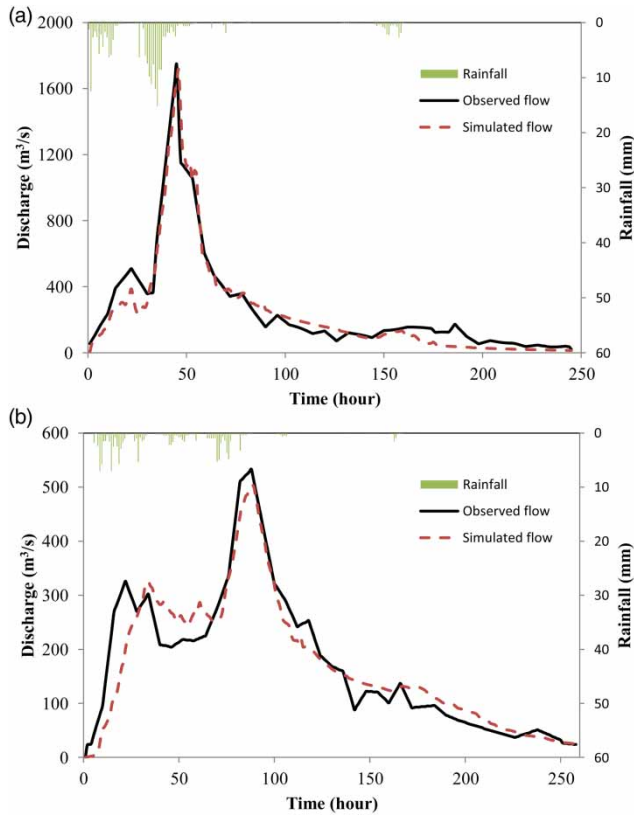
With the modified two-runoff-type TVGM, the predictability of discharge is substantially improved for low flow period as shown in Figure 4. Overall, the modified TVGM retains the mathematical simplicity in its functional form and meets the principle of parsimony (Hailegeorgis & Alfredsen 2015), with seven empirical parameters. The model directly correlates rainfall to runoff without requiring additional observations such as evaporation or soil moisture, making it attractive for hydrological response prediction, especially in data-sparse areas.

## RESULTS AND DISCUSSION

### Model calibration and verification

In this section, the modified TVGM is applied to 13 catchments in Huai River Basin. Due to limited data availability, lengths of data records for different catchments are different. For each catchment, about two-thirds of the collected rainfall-runoff data were used for model calibration while the remaining third was used for model verification. The available data in calibration and verification periods for different catchments are summarized in Table 5. As introduced previously, the parameters of the modified TVGM





**Figure 4** | Flood simulation by modified TVGM in Baillianhe Reservoir for the same two flood events as in Figure 3.

**Table 5** | Available 3-hourly data of each catchment

Hydrological station	Calibration period	Verification period
Huangchuan	2007 2008	2005
Jiangjiaji	2002 2003 2004 2005	2007
Xixian	2002 2004 2005	2007 2008
Bantai	2004 2005 2006	2007 2008
Luohe	2004 2007	2001
Mengcheng	2008	2007
Huaibin	2005 2006 2007	2008
Zhoukou	2004 2005	2000
Jieshou	2004 2007	2006 2008
Wangjiaba	2000 2002 2004 2006	2007 2008
Runheji	2000 2002 2004 2006	2007 2008
Lutaizi	2002 2004 2006 2008	2000 2008
Bengbu	2000 2002 2004 2005	2007 2008

Notes: Different hydrological stations have different lengths of available data. The time period of each year was from April 1, 8:00 to September 30, 8:00 with a time step of 3 hours.

were calibrated via the Householder least square method and genetic algorithm, as presented in Table 6.

From Table 6, it is clear that the coefficients  $g_2$  and  $g_4$  are always positive, while  $g_1$  and  $g_3$  take values around 0 (either positive or negative). This can be physically explained based on the runoff generation mechanisms described in Equations (9)–(12). If the time variant gain coefficient of quick flow, i.e.,  $G_s$  obtained from Equation (10) is positive, the generation of quick flow  $R_s$  begins according to Equation (9). If  $g_2 = 0$ ,  $G_s$  becomes a constant factor, (i.e.,  $G_s = g_1$ ) rather than a time variant factor. If  $g_2 < 0$ ,  $G_s$  becomes a time-decreasing factor, which is not reasonable since the quick flow amount increases with  $API$  or soil moisture. Therefore,  $g_2$  must be positive. On the other hand,  $g_1$  does not have the positive constraint. If  $g_1 \leq 0$ , quick flow begins to generate when the initial  $API$  is large enough, leading to  $G_s > 0$ , which is physically reasonable in arid areas. If  $g_1 > 0$ , quick flow starts instantaneously with precipitation, which is possible for humid areas. Based on a similar argument, it is straightforward to show that  $g_4$  is always positive, while  $g_3$  is negative or positive in arid or humid areas, respectively. Additionally, all flow routing parameters (i.e.,  $N$ ,  $K$ , and  $KKG$ ) are positive and  $KKG$  is found to be site-insensitive.

Two statistical metrics were used to evaluate the model performances, including the Nash and Sutcliffe efficiency (NSE) and the coefficient of water balance (CWB):

$$NSE = 1 - \frac{\sum (Y_{obs} - Y_{sim})^2}{\sum (Y_{obs} - \bar{Y}_{obs})^2} \quad (23)$$

$$CWB = \frac{\sum Y_{sim}}{\sum Y_{obs}} \quad (24)$$

where  $Y_{obs}$  and  $Y_{sim}$  represent observed and simulated flows, respectively, and  $\bar{Y}_{obs}$  is the mean value of observed flow series. According to Equations (23) and (24), the model performance is better when NSE and CWB are closer to 1. The mean NSE and CWB values for each catchment at calibration and verification periods are presented in Table 7. During calibration periods, the mean NSE of all catchments is 0.90 and the mean CWB of all catchments is 0.83. During the verification period, the mean NSE of all catchments is 0.91 and the mean CWB of all catchments is 0.94. These statistical measures indicate that the modified

**Table 6** | Calibrated parameters of the modified TVGM for each catchment

Catchment	$g_1$	$g_2$	$g_3$	$g_4$	$n$	$K$	$KKG$
Huangchuan	-0.07	0.34	0.05	0.21	3.23	1.83	0.92
Jiangjiaji	0.02	0.22	0.00	0.28	8.17	1.12	0.91
Xixian	0.00	0.26	-0.01	0.28	4.27	2.42	0.91
Bantai	-0.13	0.38	-0.09	0.38	4.25	4.21	0.91
Luohe	-0.06	0.21	-0.06	0.21	2.67	4.06	0.91
Mengcheng	0.02	0.14	0.07	0.06	7.05	2.08	0.91
Huaibin	-0.02	0.36	-0.07	0.44	7.05	3.02	0.91
Zhoukou	-0.05	0.22	-0.08	0.31	4.59	3.89	0.90
Jieshou	-0.04	0.19	-0.02	0.17	7.94	2.79	0.91
Wangjiaba	0.01	0.21	0.08	0.19	5.90	3.80	0.91
Runheji	-0.03	0.27	-0.01	0.30	8.56	4.21	0.91
Lutaizi	-0.03	0.30	-0.02	0.36	7.34	5.77	0.90
Bengbu	0.03	0.24	0.19	0.03	7.94	6.51	0.91

**Table 7** | Metrics of model performance for each catchment during calibration and verification periods

Catchment	Calibration period		Verification period	
	NSE	CWB	NSE	CWB
Huangchuan	0.95	0.82	0.92	0.83
Jiangjiaji	0.92	0.83	0.98	1.40
Xixian	0.94	0.88	0.96	0.92
Bantai	0.91	0.86	0.89	0.86
Luohe	0.91	0.83	0.99	1.28
Mengcheng	0.89	0.83	0.96	0.91
Huaibin	0.93	0.83	0.84	0.83
Zhoukou	0.94	0.76	0.96	1.24
Jieshou	0.90	0.87	0.87	0.79
Wangjiaba	0.89	0.86	0.88	0.88
Runheji	0.91	0.85	0.91	0.74
Lutaizi	0.89	0.83	0.84	0.76
Bengbu	0.82	0.74	0.88	0.84

TVGM is robust and capable of predicting the 3-hourly streamflow for catchments with reasonably good accuracy.

### Regressive regionalization analysis

After testing the modified TVGM in gauged basins, we then conduct a regionalization process to predict hydrological

processes of ungauged basins. As introduced in Razavi & Coulibaly (2012), regression is more efficient than other regionalization approaches for warm temperate regions, and is adopted in this study. To verify the applicability of regionalization, we apply the model, using parameters derived from regressive analysis rather than by calibration against actual measurements, to reproduce the previous streamflow. The general criterion for catchment selection follows that both catchments for equation derivation and verification should cover a variety of catchment size (i.e., small, medium, and large) for better representativeness and applicability of the derived equations. Specifically, in this study, eight catchments with a range of area from 5,930 to 88,630 km<sup>2</sup>, including Jiangjiaji, Xixian, Bantai, Huaibin, Zhoukou, Jieshou, Runheji, and Lutaizi were selected for derivation of regression equations, while the remaining five catchments with a range of area from 2,050 to 121,330 km<sup>2</sup>, including Huangchuan, Luohe, Mengcheng, Wangjiaba, and Bengbu were used for verification of derived regression equations (check Table 1 for more catchment area information). Due to data availability and limitation in quality, we excluded some catchments while deriving regression equations such as Mengcheng (highly impacted by human activities) and Bengbu (with very large area and strong heterogeneous land surface conditions). The regression approach can be repeated with different

combinations of catchments for calibration and validation. Under different combination cases, the empirical coefficients of the proposed model, being site-specific, will change correspondingly whereas the impact on the overall model predictability is insignificant.

In general, flow dynamics of a catchment is closely related to dominant climatic conditions and catchment characteristics (Castiglioni et al. 2010; Gibbs et al. 2012). Located in the same basin (Huai River Basin), the selected catchments should have similar climatic conditions. Hence we focus on the differences caused by underlying surface characteristics, including soil moisture, soil properties, soil structure, vegetation conditions, topography, etc. Four regression equations on runoff generation parameters (i.e.,  $g_1, g_2, g_3, g_4$ ) are derived from underlying surface characteristics via SPSS statistics software with a significance level of 0.05 as shown in Equations (25)–(28). The  $R^2$  values for Equations (25)–(28) are 0.95, 0.94, 0.93, and 0.94, respectively:

$$-g_1/g_2 = -0.03top - 0.42h + 3.87W_f/(W_s - W_w) - 5.29W_s/(W_s - W_w) + 5.63 \quad (25)$$

$$g_1 + g_2W_s = -0.42f_1 - 0.18f_2 + 0.20 \quad (26)$$

$$g_3 + g_4W_s = -0.34f_1 + 0.26f_3 + 0.07 \quad (27)$$

$$g_3 = 0.79g_2 - 0.64g_4 - 0.48AW + 0.61 \quad (28)$$

where  $top$  is the topographic wetness index (m) related to soil attributes proposed by Beven & Kirkby (1979) ( $top = \ln \frac{a}{\tan b}$  with  $a$  denoting the upstream contributing area per unit contour length and  $\tan b$  denoting the local slope);  $h$  is the depth of soil layer (m);  $W_f, W_s, W_w$  are field capacity (%), saturated capacity (%), and wilting point (%) of soil, respectively;  $f_1, f_2, f_3$  are the fractions of grassland (RNGE), agricultural dry land (AGRR), and paddy field (RICE), respectively;  $AW$  is the available water capacity (in/ft), which indicates the inches of water needed to refill a foot of soil to field capacity.

Since a quick flow gain coefficient  $G_s$  ( $G_s \geq 0$ ) can be written as  $G_s = g_1 + g_2API = g_2(g_1/g_2 + API)$ , and  $g_2$  is always positive, we have  $API \geq -g_1/g_2$ . In Equation (25),

$-g_1/g_2$ , as a critical indicator of quick flow generation, is found to be dependent on land surface characteristics including topographic index and soil properties. When the soil moisture is saturated, the time variant gain coefficients for quick flow and baseflow can be represented by  $G_s = g_1 + g_2W_s$  and  $G_g = g_3 + g_4W_s$ . The two flow generation coefficients are found to be dependent on different land use types, as shown in Equations (26) and (27). Due to the constraints of available water capacity, the coefficients of baseflow and quickflow gain factors are interrelated as in Equation (28). The left-hand side (LHS) terms of Equations (25)–(28) are calculated with calibrated parameters, while the right-hand side (RHS) terms are calculated from watershed descriptors. The comparisons between the calibrated parameters (LHS) and the derived ones from watershed descriptors (RHS) are shown in Figure 5, where  $R^2$  values are found greater than 0.93 for each comparison.

For the three flow routing parameters, we only focus on regionalization of  $n$  and  $K$ , since the model predictions are insensitive to  $KKG$ . Based on SPSS statistical analysis, two regression equations on  $n, K$  with  $R^2$  values of 0.82 and 0.79, respectively, are derived as below:

$$nK = 0.0004A + 12.288 \quad (29)$$

$$K = 4.712 \exp(-19.23P\alpha/d) \quad (30)$$

where  $A$  is the catchment area ( $\text{km}^2$ ),  $P$  is the long-term average rainfall (mm),  $\alpha$  is the mainstream slope, and  $d$  is the distance between the rainfall centroid and the catchment outlet (km). The relationships between  $nK$  and  $A$  as well as between  $K$  and  $p\alpha/d$  are plotted in Figure 6. These two regressive equations have reasonable goodness-of-fit in comparison with the calibrated values with the  $R^2$  of around 0.80 and also physically reasonable. Since  $n$  and  $K$  refer to the number of linear reservoirs and storage-discharge coefficient of each reservoir, respectively, the multiplication  $nK$  is equivalent to the watershed concentration time, which is related to the catchment area. As well,  $K$  as a parameter of storage-discharge capacity can influence the deformation of flood curve and should be negatively related to rainfall intensity and positively related

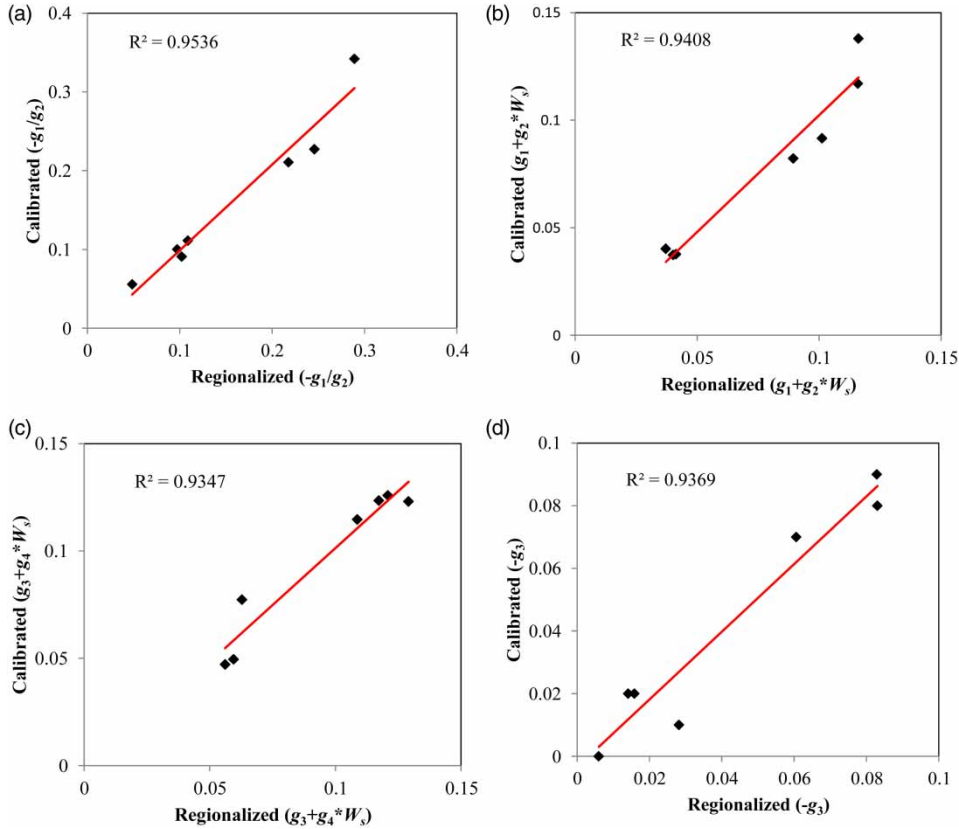


Figure 5 | Comparison between calibrated variables and regionalized variables, i.e., variables calculated by regression equations.

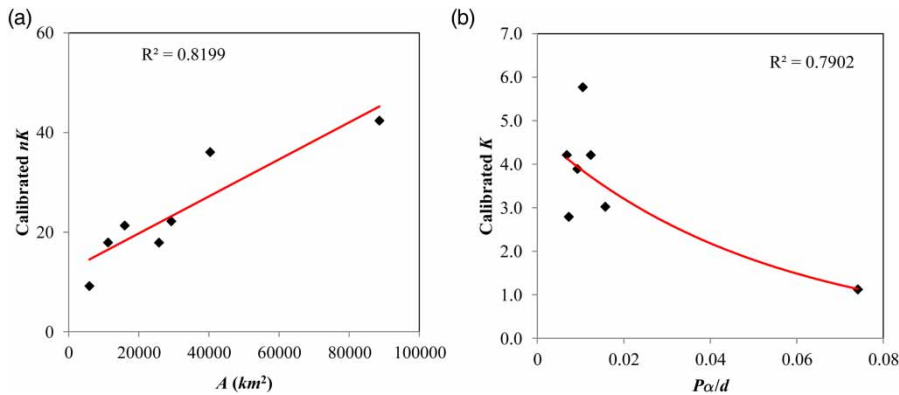


Figure 6 | The relationship between runoff routing variables ( $n$ ,  $K$ ) and watershed characteristics, including watershed area ( $A$ ), average precipitation ( $P$ ), main stream slope ( $\alpha$ ) and distance between the rainfall centroid and the catchment outlet ( $d$ ).

to the distance between rainfall centroid and catchment outlet.

With the six regression equations, i.e., Equations (25)–(30), we can calculate the model parameters for the

remaining five catchments, Huangchuan, Luohe, Wangjiaba, Mengcheng, and Bengbu based on their own catchment descriptors. Then the modified TVGM was applied to reproduce all the observed streamflows at the outlet of each

catchment, namely, streamflows of 2005, 2007, 2008 for Huangchuan station, streamflows of 2001, 2004, 2007 for Luohe station, streamflows of 2007, 2008 for Mengcheng station, streamflows of 2000, 2002, 2004, 2006, 2007, 2008 for Wangjiaba station, and streamflows of 2000, 2002, 2004, 2005, 2007, 2008 for Bengbu station. Note that only data from April 1, 8:00 to September 30, 8:00 (flood season) for each year are analyzed in this paper. It is noteworthy that the selected study area is subject to human interference such as the presence of sluice gates and dams, the original hydrographs having been changed over the course of investigation. The reconstruction of natural streamflow presents as challenging due to the lack of sluice operation data, especially for low-flow periods. Given that sluice gates are usually open during the flood season, the rainfall–runoff data of the flood season were selected for analysis to weaken the impacts of human activities. The regionalized model parameters and model performance indicators (i.e., NSE and CWB) of the five catchments are presented in Table 8. The simulated 3-hourly streamflows by regionalization are compared with observed streamflows in Figure 7.

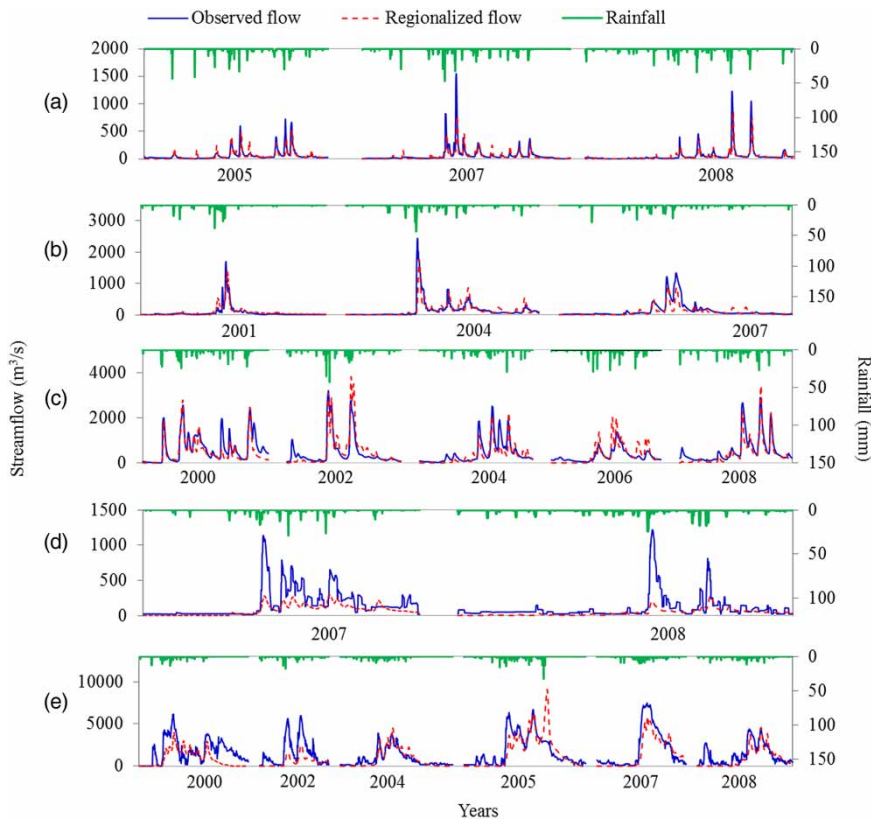
The streamflows of Luohe and Wangjiaba were best simulated among the five catchments with mean NSE of 0.83 and 0.89, respectively, and mean CWB of 0.93 and 0.92, respectively. The two catchments are medium size catchments with an area of 12,150 km<sup>2</sup> and 30,630 km<sup>2</sup>, covering about 10% and 40% of the total study area, respectively. Thus the catchment characteristics are similar to average characteristics of the study area and can be captured by the above derived regression equations with reasonable accuracy. In contrast, two extreme examples, including Huangchuan catchment, i.e., the smallest sub-catchment with an area of only 2,050 km<sup>2</sup> and Bengbu catchment, i.e., the largest

sub-catchment with an area of 121,330 km<sup>2</sup> are analyzed (see Figures 1 and 2). The streamflows of Huangchuan and Bengbu are predicted with mean NSE of 0.70 and 0.76, respectively, and mean CWB of 0.79 and 0.69, respectively. Compared with Luohe and Wangjiaba catchments, the streamflow of Huangchuan catchment is predicted with reduced accuracy since it is practically difficult to represent the land surface characteristics at the upstream area of such a small catchment by average regionalization equations. As for Bengbu, given that it has the largest drainage area with various underlying surfaces and numerous sluice gates and dams, the modeling of hydrological balance in the regionalized sense is computationally challenging. Lastly, the prediction of streamflows of Mengcheng catchment is of least accuracy with mean NSE of 0.58 and mean CWB of 0.45. The primary reason may be that Mengcheng is located at the boundary of the study area (see Figures 1 and 2) so that the average regionalization equations are less representative, especially for the land surface characteristics across the watershed boundaries. Also, observed flows are severely influenced by sluice gates and dams. From Figure 7(d), the hydrographs at Mengcheng hydrological station are often attenuated as a straight line parallel to time axis, which implies that the outflow of Mengcheng is manipulated by human control to maintain a constant discharge value.

In addition, the equifinality phenomenon (Beven & Freer 2001; Todini 2007; Zhang et al. 2012; Hailegeorgis & Alfredsen 2015), i.e., same model performances with different model parameters, inevitably affects the results of parameter calibration. By analyzing the model parameters, we determine the ranges of calibrated parameters with the most probable values being site-specific so that the uncertainty resulting from equifinality can be reduced. Overall,

**Table 8** | Model parameters and evaluation of regionalization

Catchment	Regionalized model parameters						Model evaluation	
	$\xi_1$	$\xi_2$	$\xi_3$	$\xi_4$	$n$	$K$	Mean NSE	Mean CWB
Huangchuan	0.02	0.22	0.10	0.11	14.75	0.80	0.70	0.79
Luohe	-0.04	0.21	0	0.13	7.69	2.06	0.83	0.93
Mengcheng	0.01	0.04	0.02	0.1	3.69	4.65	0.58	0.45
Wangjiaba	-0.03	0.35	0.01	0.3	5.97	3.89	0.89	0.92
Bengbu	-0.03	0.29	-0.14	0.68	13.81	4.31	0.76	0.69



**Figure 7** | Comparison of observed flow and regionalized flow in five catchments: (a) Huangchuan, (b) Luohe, (c) Wangjiaba, (d) Mengcheng, and (e) Bengbu.

regression is an effective regionalization approach for predictions of streamflow and other hydrological responses in ungauged regions by transferring the relationships between model parameters and catchment characteristics that are established in gauged basins. Due to limited data availability, the land use and cover change of this study area were not considered in this article, but will be pursued in a future study.

## CONCLUSION

This paper presents a regressive model incorporating regionalized watershed information for streamflow predictions in ungauged basins based on a hydrological system approach (i.e., the modified TVGM). This approach converts rainfall to streamflow through physically based mathematical transformations without information of evaporation or soil moisture. Also, the model has parsimonious physical parameters that

can be derived from watershed descriptors (such as the underlying surface properties and precipitation characteristics) according to derived regressive equations. The regressive model was applied to predict the streamflow of five ungauged catchments in Huai River Basin with reasonable accuracy, demonstrating its effectiveness for hydrological prediction and water resources management in data-sparse areas.

In a future study, we will incorporate more watershed descriptors such as climatic factors while establishing regressive equations, and test the regressive model robustness in more watersheds with different characteristics. In addition, physical mechanisms should be carefully investigated not only during the selection of correlative factors but also during the determination of linear or nonlinear regression relations. The model can be further improved with more available measurement datasets as well as regional geographic information. In the context of rapid urbanization within major watersheds in China, future model development using regression analysis by

incorporation of, for example, anthropogenic factors will be critical in understanding the evolution of catchments' hydrological responses to potential landscape modification scenarios. The numerical predictions of the improved regionalization model will also shed new light onto the modified physics of hydrological cycle under emergent climatic patterns, and provide useful guidelines for sustainable landscape planning in a developing China.

## ACKNOWLEDGEMENTS

This study was supported by the National Natural Science Foundation of China (No. 41571028 and No. 51279139). The authors J. Song and Z. H. Wang are also supported by the US National Science Foundation under grant number CBET-1435881.

## REFERENCES

- Abdulla, F. A. & Lettenmaier, D. P. 1997 Development of regional parameter estimation equations for a macroscale hydrologic model. *J. Hydrol.* **197**, 230–257.
- Ahsan, M. & O'Connor, K. M. 1994 A simple non-linear rainfall-runoff model with a variable gain factor. *J. Hydrol.* **155**, 151–183.
- Beven, K. & Freer, J. 2001 Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *J. Hydrol.* **249**, 11–29.
- Beven, K. J. 2011 *Rainfall-runoff Modelling: The Primer*. John Wiley & Sons, Chichester.
- Beven, K. J. & Kirkby, M. J. 1979 A physically based, variable contributing area model of basin hydrology. *Hydrol. Sci. Bull.* **24**, 43–69.
- Blöschl, G. & Sivapalan, M. 1995 Scale issues in hydrological modelling: a review. *Hydrol. Process.* **9**, 251–290.
- Businger, P. & Golub, G. H. 1965 Linear least squares solutions by householder transformations. *Numer. Math.* **7**, 269–276.
- Carassale, L. & Kareem, A. 2009 Modeling nonlinear systems by volterra series. *J. Eng. Mech.* **136**, 801–818.
- Castellarin, A., Camorani, G. & Brath, A. 2007 Predicting annual and long-term flow-duration curves in ungauged basins. *Adv. Water Res.* **30**, 937–953.
- Castiglioni, S., Lombardi, L., Toth, E., Castellarin, A. & Montanari, A. 2010 Calibration of rainfall-runoff models in ungauged basins: a regional maximum likelihood approach. *Adv. Water Res.* **33**, 1235–1242.
- Chapman, T. 1999 A comparison of algorithms for stream flow recession and baseflow separation. *Hydrol. Process.* **13**, 701–714.
- Chen, G., Xie, P. & Xiong, L. 2009 Spatial and temporal prediction of flash floods in ungauged area based on GIS. In: *Advances in Water Resources and Hydraulic Engineering* (C. Zhang & H. Tang, eds). Springer, Berlin, pp. 113–118.
- Chen, J., Yu, Z., Zhu, Y. & Yang, C. 2011 Relationship between land use and evapotranspiration – a case study of the Wudaogou area in Huaihe river basin. *Procedia Environ. Sci.* **10**, 491–498.
- Cibin, R., Athira, P., Sudheer, K. & Chaubey, I. 2014 Application of distributed hydrological models for predictions in ungauged basins: a method to quantify predictive uncertainty. *Hydrol. Process.* **28**, 2033–2045.
- Cleveland, T. G., Thompson, D. B., Fang, X. & He, X. 2008 Synthesis of unit hydrographs from a digital elevation model. *J. Irrig. Drain. Eng.* **134**, 212–221.
- Eckhardt, K. 2005 How to construct recursive digital filters for baseflow separation. *Hydrol. Process.* **19**, 507–515.
- Gibbs, M., Maier, H. & Dandy, G. 2012 A generic framework for regression regionalization in ungauged catchments. *Environ. Modell. Softw.* **27**, 1–14.
- Hailegeorgis, T. T. & Alfredsen, K. 2015 Comparative evaluation of performances of different conceptualisations of distributed HBV runoff response routines for prediction of hourly streamflow in boreal mountainous catchments. *Hydrol. Res.* **46** (4), 607–628.
- Jin, X., Xu, C. Y., Zhang, Q. & Chen, Y. D. 2009 Regionalization study of a conceptual hydrological model in Dongjiang basin, south China. *Quatern. Int.* **208** (1), 129–137.
- Kizza, M., Guerrero, J.-L., Rodhe, A., Xu, C.-Y. & Ntale, H. K. 2013 Modelling catchment inflows into Lake Victoria: regionalisation of the parameters of a conceptual water balance model. *Hydrol. Res.* **44** (5), 789–808.
- Li, M., Shao, Q., Zhang, L. & Chiew, F. H. S. 2010 A new regionalization approach and its application to predict flow duration curve in ungauged basins. *J. Hydrol.* **389**, 137–145.
- Makungo, R., Odiyo, J., Ndiritu, J. & Mwaka, B. 2010 Rainfall-runoff modelling approach for ungauged catchments: a case study of Nzhelele River sub-quaternary catchment. *Phys. Chem. Earth PT A/B/C* **35**, 596–607.
- Mosley, M. 1981 Delimitation of New Zealand hydrologic regions. *J. Hydrol.* **49**, 173–192.
- Napiórkowski, J. J. 1992 Comment on 'identification of a constrained nonlinear hydrological system described by Volterra functional series' by J. Xia. *Water Resour. Res.* **28**, 2547–2549.
- Nash, J. 1957 The form of the instantaneous unit hydrograph. *Comptes Rendus et Rapports Assemblee Generale de Toronto* **3**, 114–121.
- Oudin, L., Andreassian, V., Perrin, C., Michel, C. & Le Moine, N. 2008 Spatial proximity, physical similarity, regression and ungauged catchments: a comparison of regionalization approaches based on 913 French catchments. *Water Resour. Res.* **44**, W03413.
- Parada, L. M. & Liang, X. 2010 A novel approach to infer streamflow signals for ungauged basins. *Adv. Water Res.* **33**, 372–386.
- Patil, S. & Stieglitz, M. 2012 Controls on hydrologic similarity: role of nearby gauged catchments for prediction at an ungauged catchment. *Hydrol. Earth Syst. Sci.* **16**, 551–562.

- Pedersen, J. T., Peters, J. C. & Helweg, O. J. 1980 Hydrographs by single linear reservoir model. *J. Hydrol.* **106**, 837–852.
- Post, D. A. & Jakeman, A. J. 1999 Predicting the daily streamflow of ungauged catchments in SE Australia by regionalising the parameters of a lumped conceptual rainfall-runoff model. *Ecol. Model.* **123**, 91–104.
- Purcell, P. J. 2006 Physical analog of the linear reservoir. *J. Hydrol. Eng.* **11**, 184–187.
- Rabunal, J., Puertas, J., Suarez, J. & Rivero, D. 2007 Determination of the unit hydrograph of a typical urban basin using genetic programming and artificial neural networks. *Hydrol. Process.* **21**, 476–485.
- Razavi, T. & Coulibaly, P. 2012 Streamflow prediction in ungauged basins: review of regionalization methods. *J. Hydrol. Eng.* **18**, 958–975.
- Reed, D. 1999 *Procedures for Flood Frequency Estimation Volume 3: Statistical Procedures for Flood Frequency Estimation*. Flood Estimation Handbook. Centre for Ecology and Hydrology, Wallingford, UK.
- Seibert, J. 1999 Regionalisation of parameters for a conceptual rainfall-runoff model. *Agric. For. Meteorol.* **98**, 279–293.
- Sherman, L. K. 1932 Streamflow from rainfall by the unit-graph method. *Eng. News Record* **108**, 501–505.
- Singh, V. P. 1988 *Hydrologic Systems: Rainfall-runoff Modeling*. Prentice Hall, Englewood Cliffs, NJ, USA.
- Sivapalan, M., Takeuchi, K., Franks, S., Gupta, V., Karambiri, H., Lakshmi, V., Liang, X., McDonnell, J., Mendiondo, E. & O'Connell, P. 2003 IAHS Decade on predictions in ungauged basins (PUB), 2003–2012: shaping an exciting future for the hydrological sciences. *Hydrolog. Sci. J.* **48**, 857–880.
- Song, J. & Wang, Z. H. 2015a Interfacing the urban land-atmosphere system through coupled urban canopy and atmospheric models. *Bound-Lay Meteorol.* **154**, 427–448.
- Song, J. & Wang, Z. H. 2015b Impacts of mesic and xeric urban vegetation on outdoor thermal comfort and microclimate in Phoenix, AZ. *Build Environ.* **5**, 918–920.
- Stainton, R. T. & Metcalfe, R. A. 2007 Characterisation and classification of flow regimes of natural rivers in Ontario to support the identification of potential reference basins. Waterpower Project Science Transfer Rep. 7.0, Ministry of Natural Resources, 90, Ottawa.
- Todini, E. 2007 Hydrological catchment modelling: past, present and future. *Hydrol. Earth Syst. Sci.* **11**, 468–482.
- Vandewiele, G. & Elias, A. 1995 Monthly water balance of ungauged catchments obtained by geographical regionalization. *J. Hydrol.* **170**, 277–291.
- Xia, J. 1991 Identification of a constrained nonlinear hydrological system described by Volterra functional series. *Water Resour. Res.* **27**, 2415–2420.
- Xia, J., O'Connor, K., Kachroo, R. & Liang, G. 1997 A non-linear perturbation model considering catchment wetness and its application in river flow forecasting. *J. Hydrol.* **200**, 164–178.
- Xia, J., Wang, G., Tan, G., Ye, A. & Huang, G. H. 2005 Development of distributed time-variant gain model for nonlinear hydrological systems. *Science in China Series D: Earth Sciences* **48**, 713–723.
- Xu, C. Y. 1999 Estimation of parameters of a conceptual water balance model for ungauged catchments. *Water Resour. Manage.* **13** (5), 353–368.
- Xu, C.-Y. 2003 Testing the transferability of regression equations derived from small sub-catchments to large area in central Sweden. *Hydrol. Earth Syst. Sci.* **7** (3), 317–324.
- Young, P. & Beven, K. 1991 Computation of the instantaneous unit hydrograph and identifiable component flows with application to two small upland catchments – comment. *J. Hydrol.* **129**, 389–396.
- Zhang, Y., Xia, J., Liang, T. & Shao, Q. 2010 Impact of water projects on river flow regimes and water quality in Huai River Basin. *Water Resour. Manage.* **24**, 889–908.
- Zhang, L., Song, X., Xia, J., Yuan, R., Zhang, Y., Liu, X. & Han, D. 2011 Major element chemistry of the Huai River Basin, China. *Appl. Geochem.* **26**, 293–300.
- Zhang, D., Zhang, L., Guan, Y., Chen, X. & Chen, X. 2012 Sensitivity analysis of Xinanjiang rainfall-runoff model parameters: a case study in Lianghui, Zhejiang Province, China. *Hydrol. Res.* **43**, 123–134.

First received 7 August 2015; accepted in revised form 15 November 2015. Available online 29 December 2015