Supplementary materials for

Land use changes is driving factor of runoff yield: Evidence from a remote sensing-based runoff generation simulation

Chaowei Xu [1], Hao Fu [1,*] , Jiashuai Yang [1] and Chan Gao[2]

1 College of Urban and Environmental Sciences, Peking University, No. 5 Yiheyuan Road,Haidian District, Beijing 100871, China

2 Univ. Bordeaux, UMR EPOC CNRS 5805, F-33400 Talence, France

*Correspondence: fuhao@pku.edu.cn

Figures:

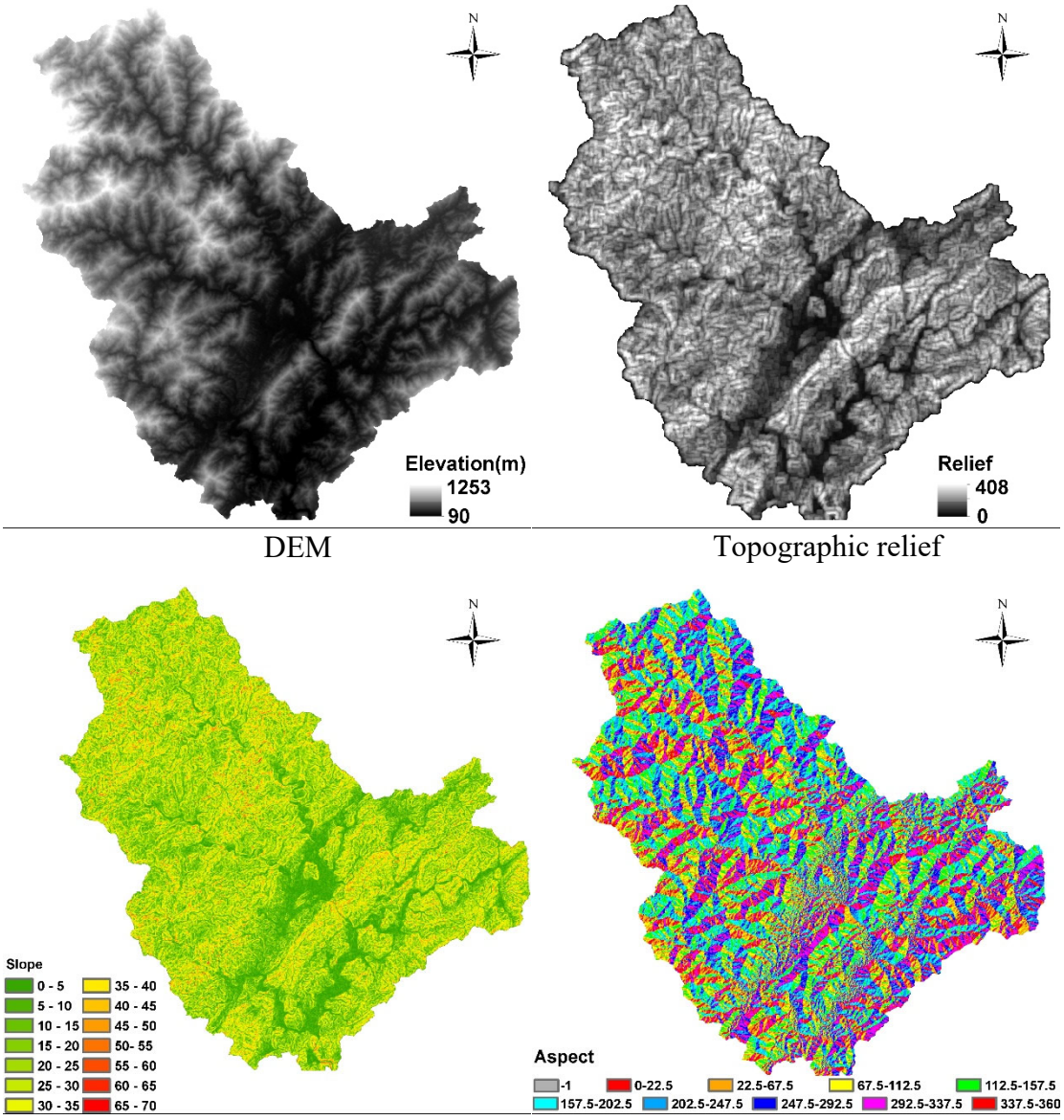


Figure S1. Topographic information of the study area.

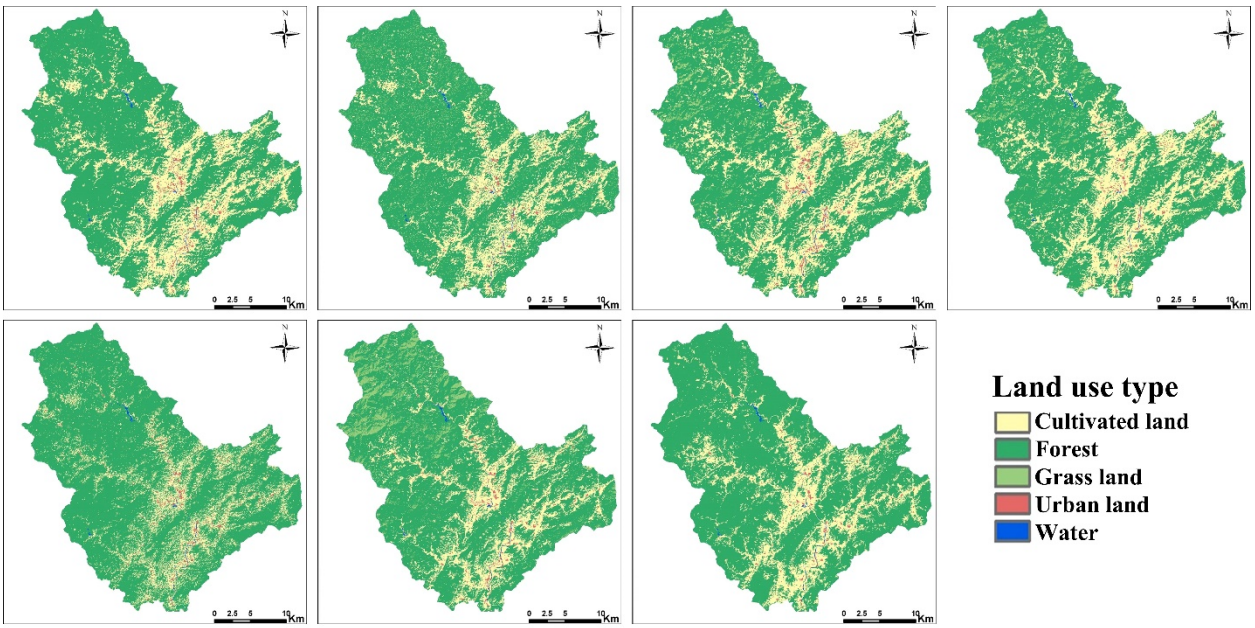
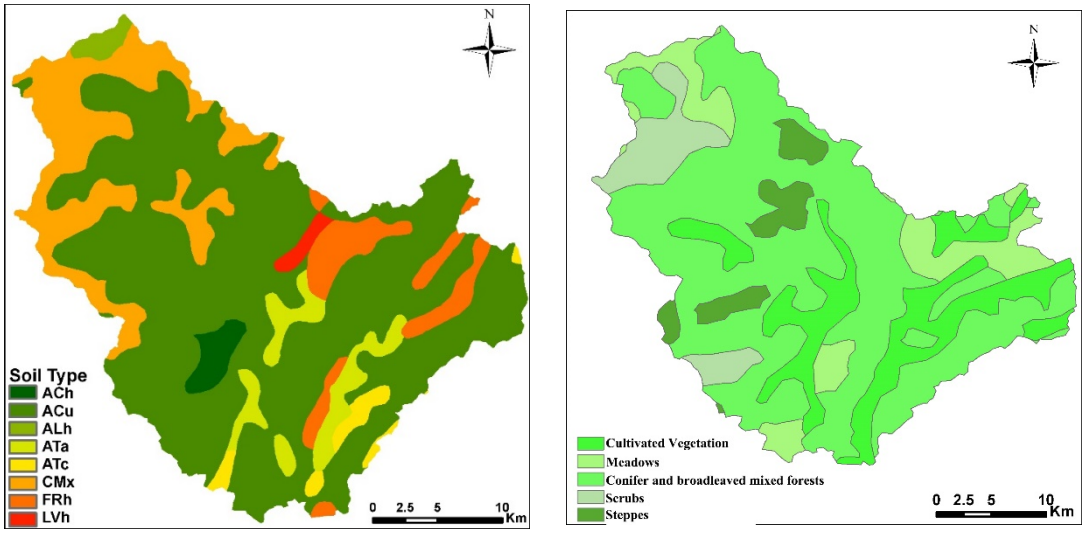


Figure S2. Land use of study area.



(a) **(b)**

Figure S3. The soil and vegetation of the study area. (a) Soil types (b) vegetation types.

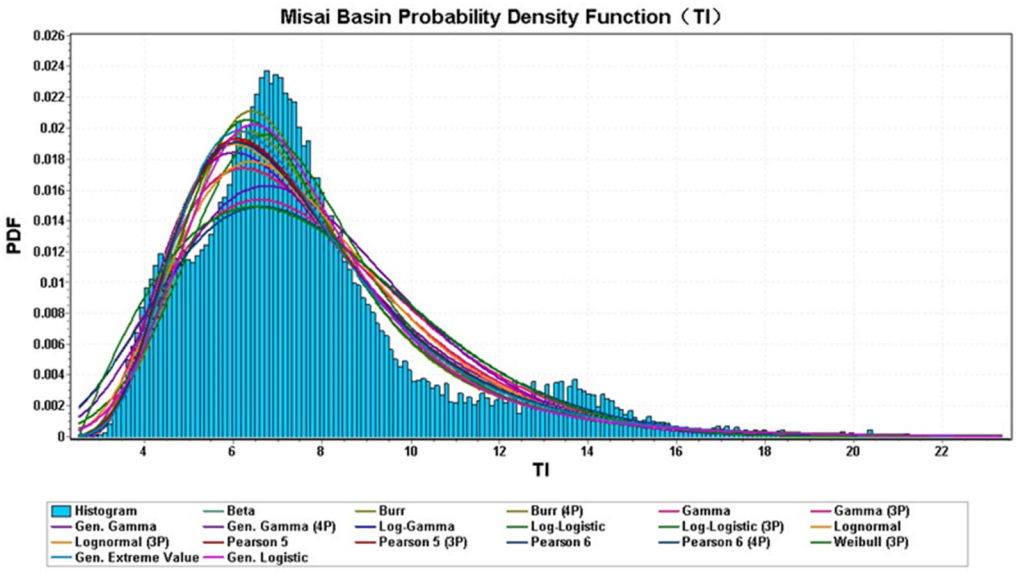


Figure S4. Different fitting functions of the probability density of topographic index.

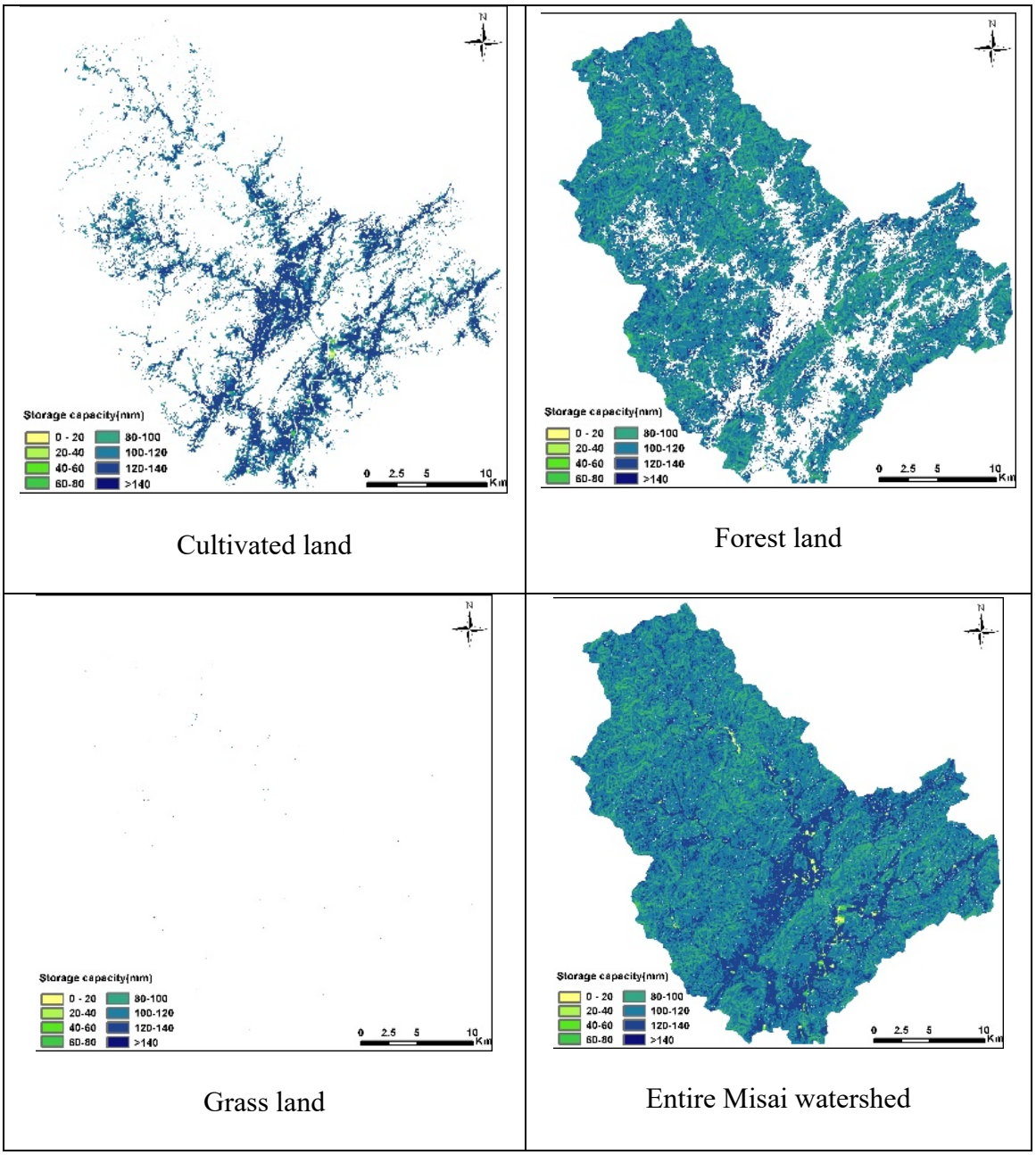


Figure S5. Spatial distribution of water storage capacity under different land use (The water storage capacity of Water and Urban land is 0, not shown).

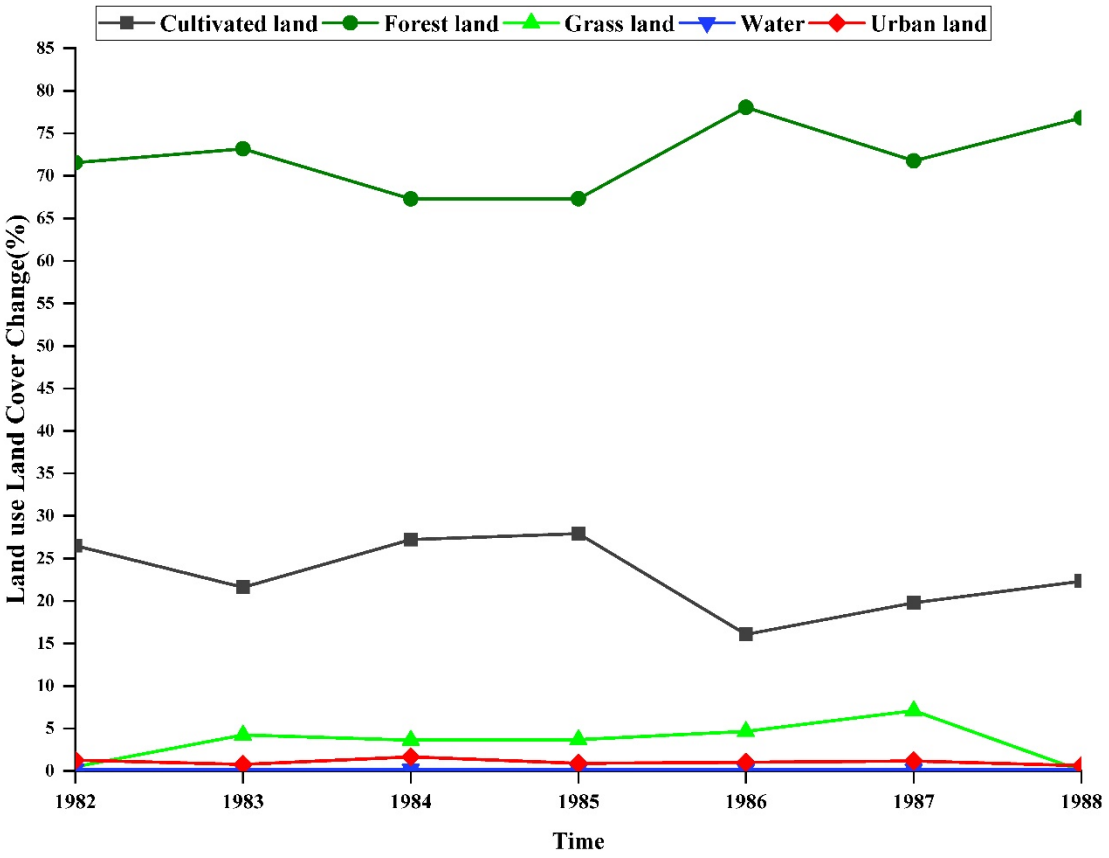


Figure S6. Land use land cover dynamics in Misai basin.

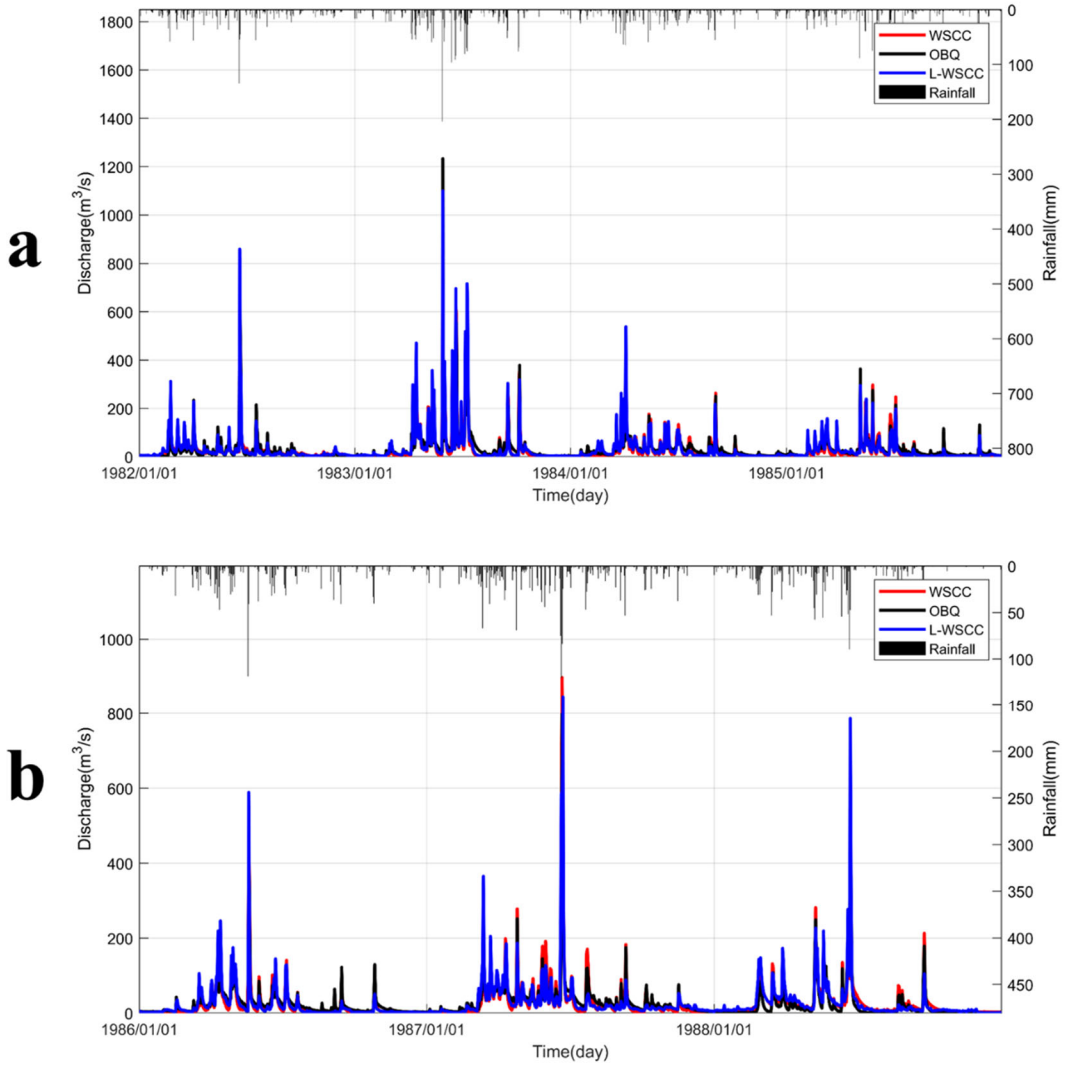


Figure S7. The observed and simulated streamflow hydrographs (daily scale) during: a: calibration and b: validation period.

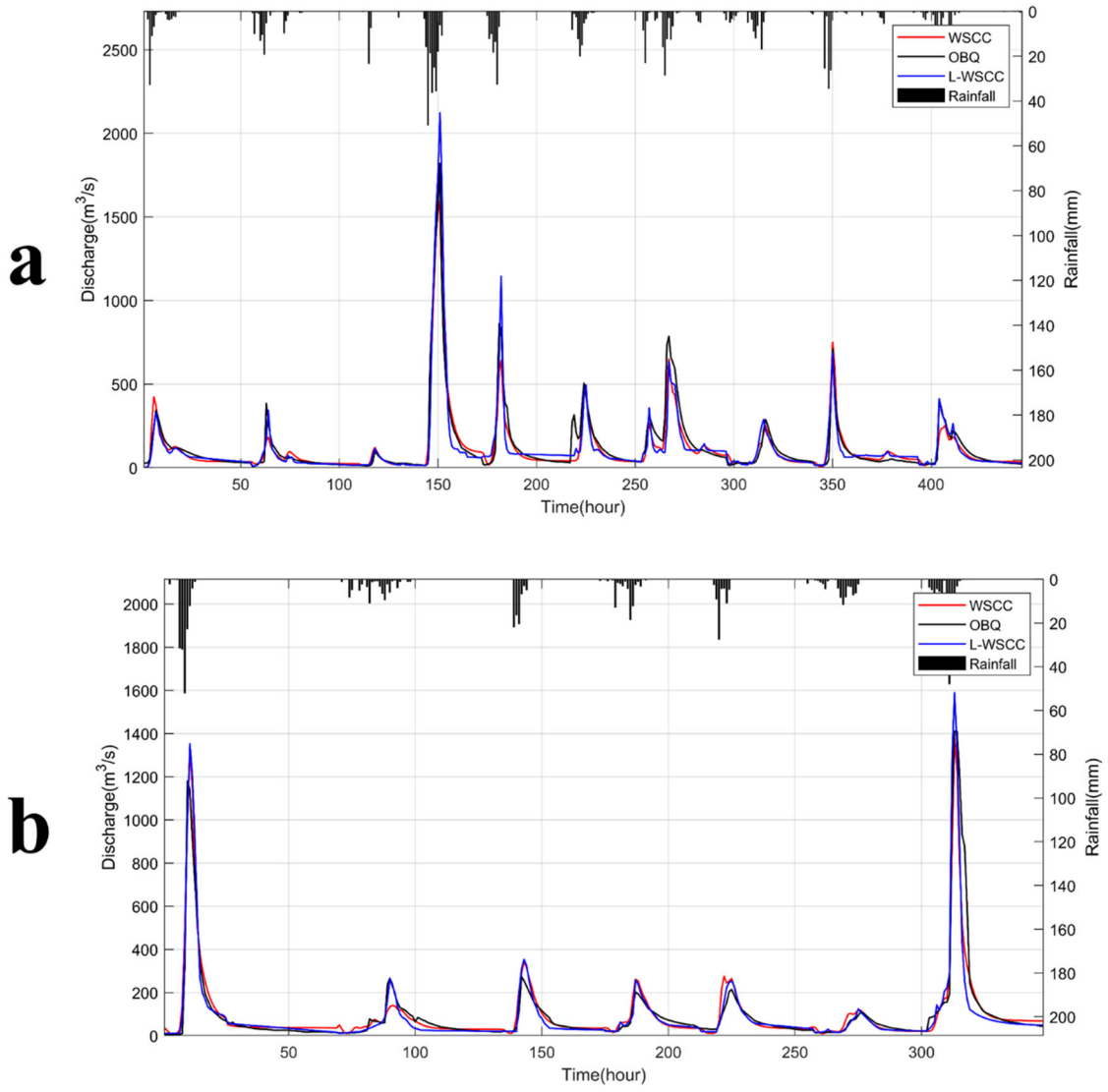


Figure S8. Comparison between observed and simulated discharge hydrographs (hourly scale) during: a: calibration and b: validation period.

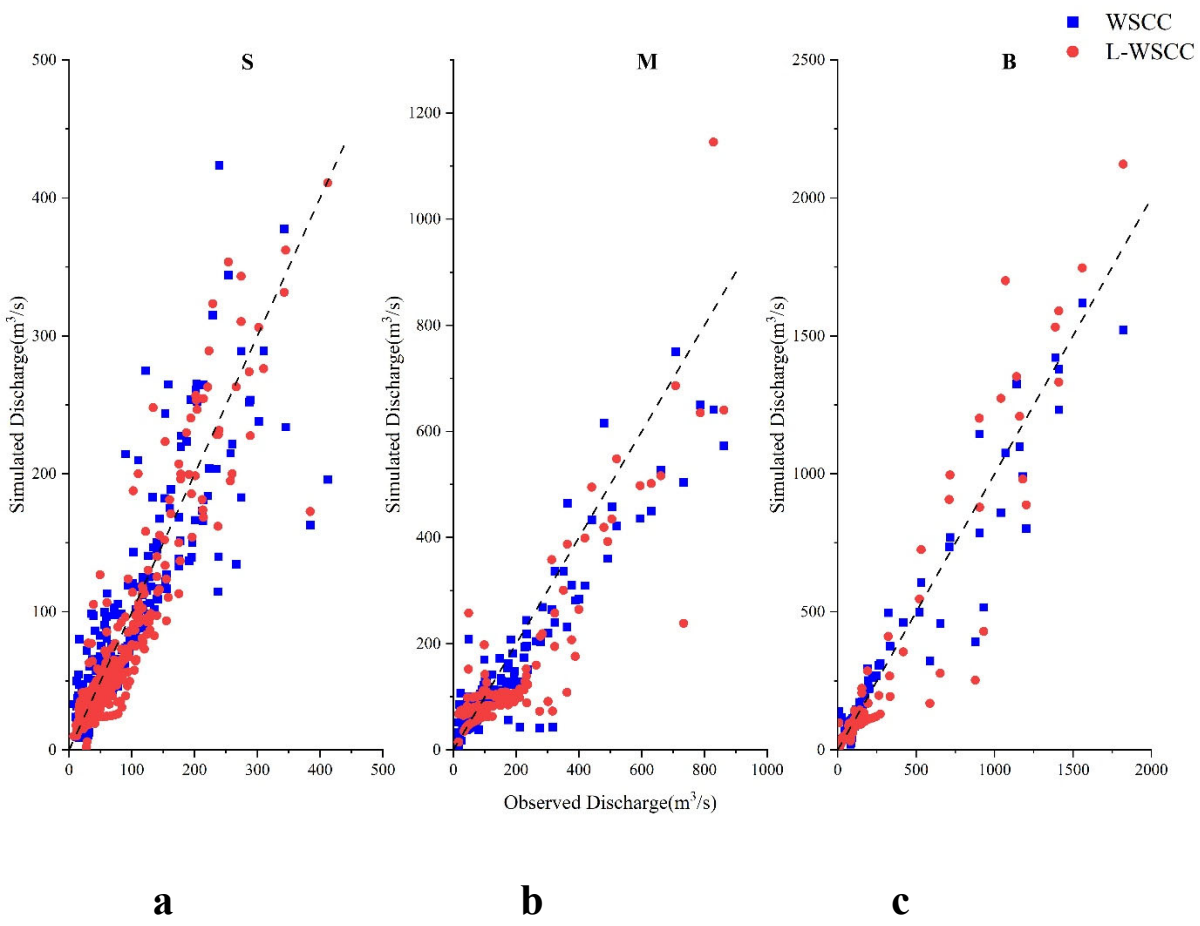


Figure S9. Simulated and observed discharge under different flood types (all calibration and validation period) (a: big floods, b: medium floods, c: small floods).

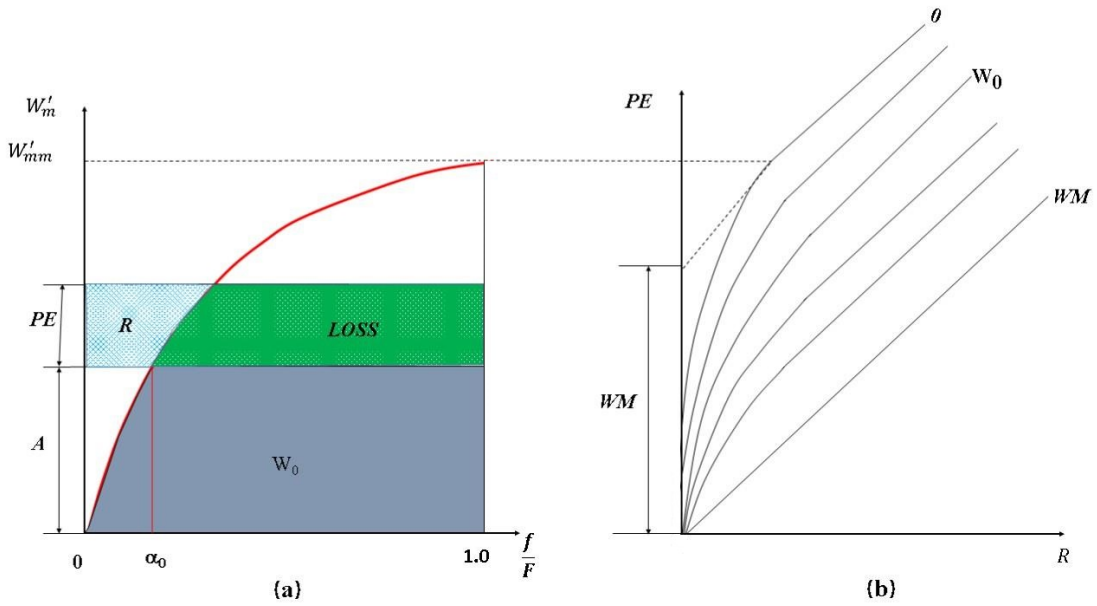
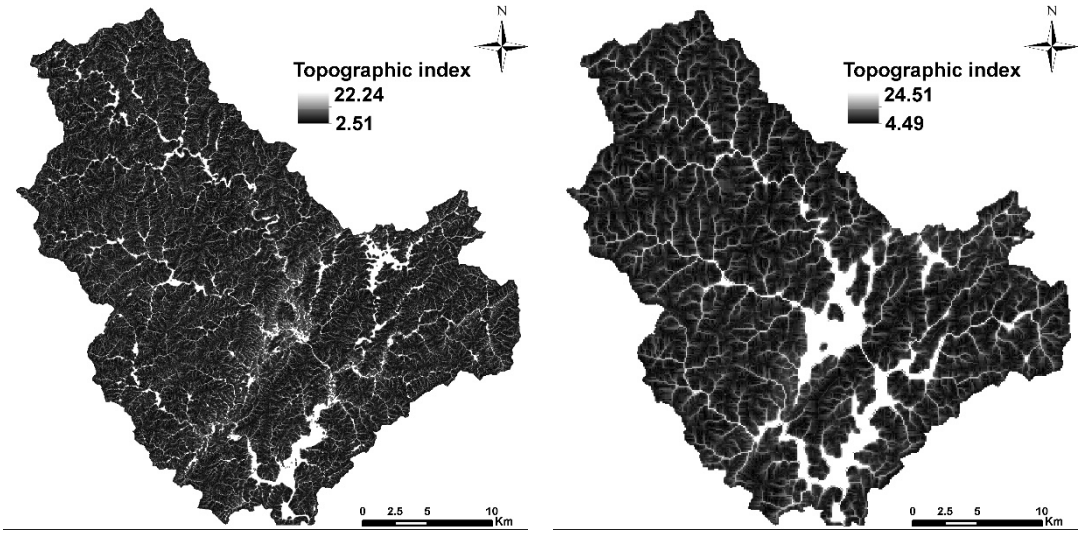


Figure S10. (**a**) The water storage capacity curve; (**b**) Rainfall-runoff relationship.



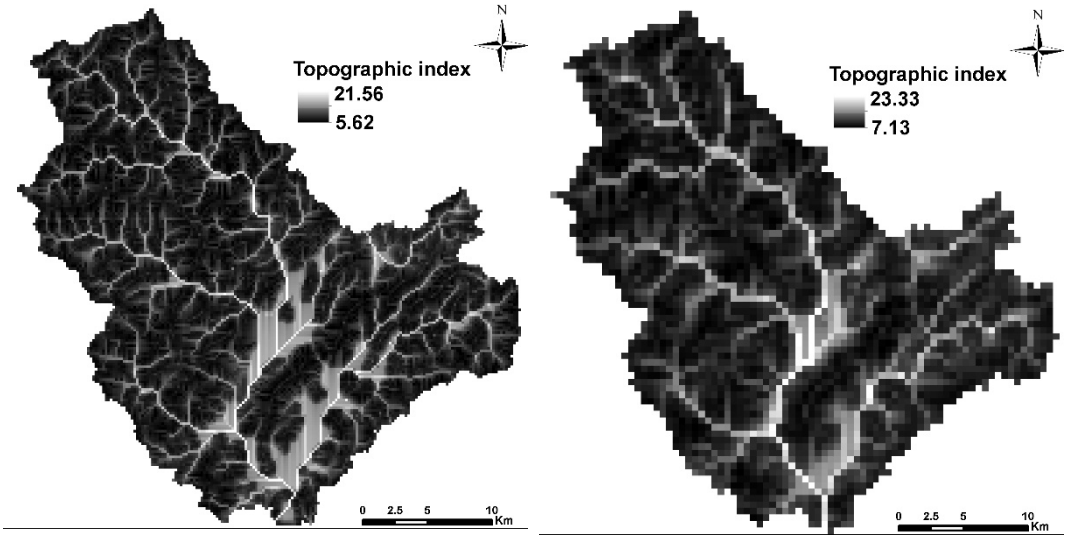


Figure S11. Topographic index distribution under four different scales in Misai basin (30m,100m,200m,500m).

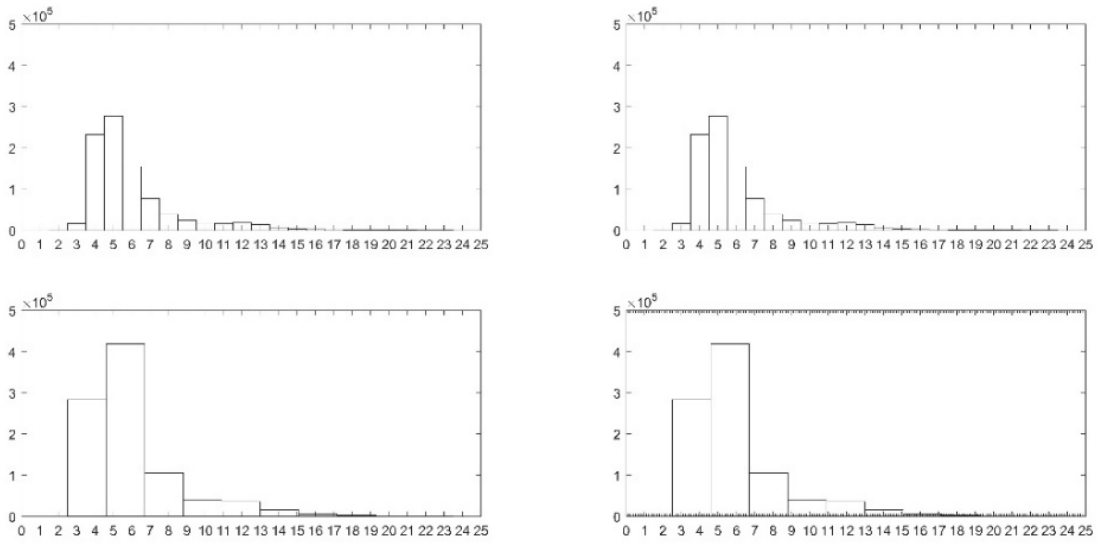


Figure S12. Topographic index frequency distribution histogram under four different scales in the Misai basin (30m,100m,200m,500m).

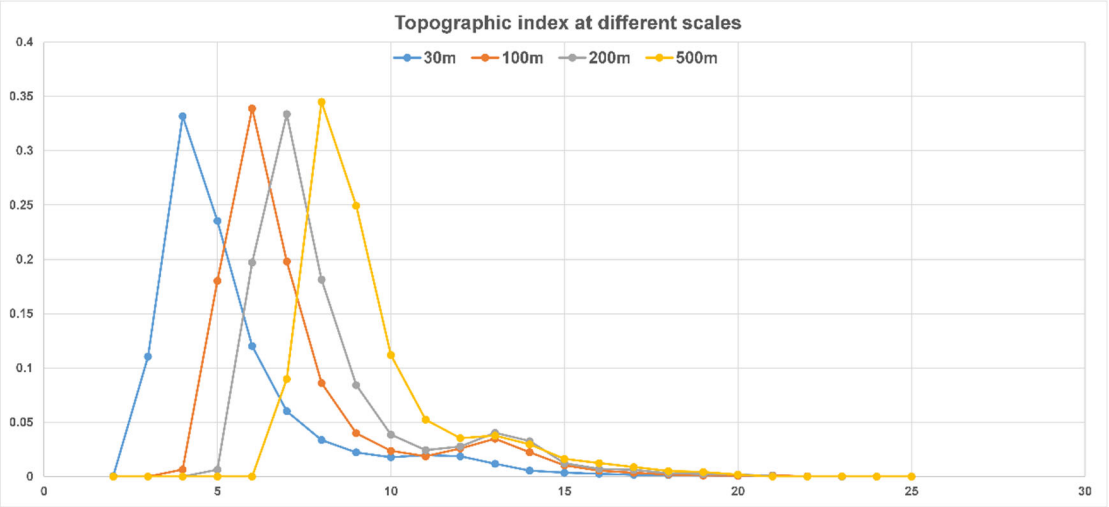


Figure S13. Probability density of topographic index at different scales.

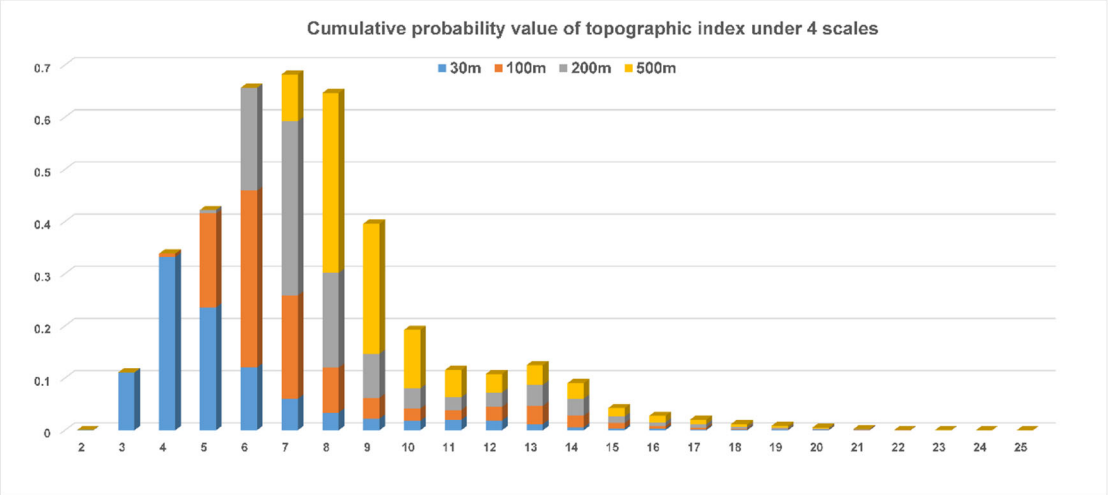


Figure S14. Probability density of topographic index (the sum of the 4 scales).

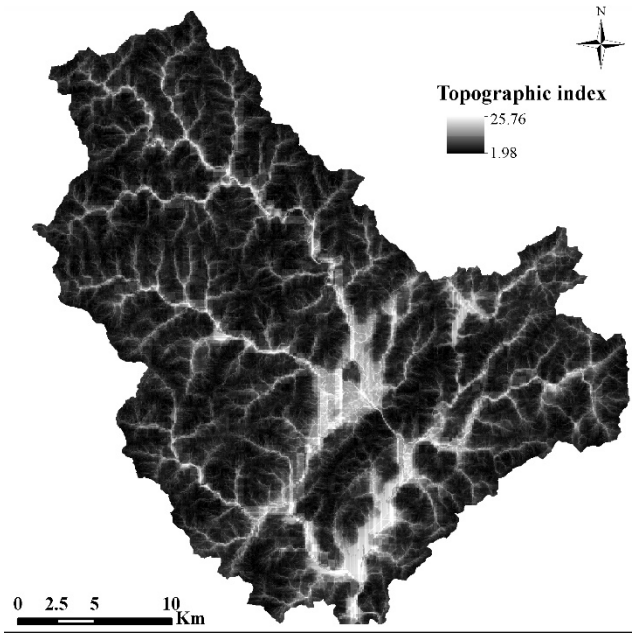
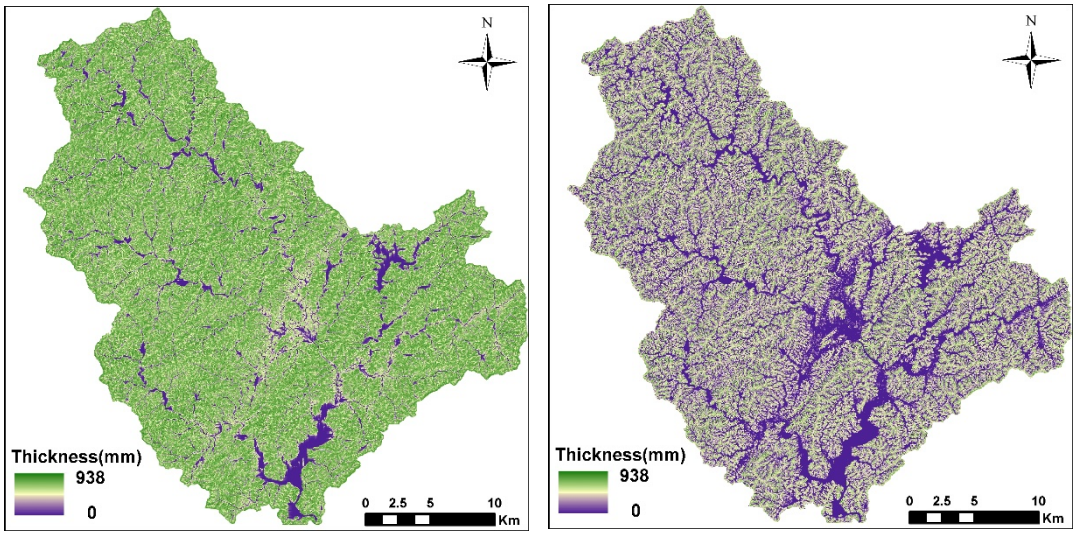
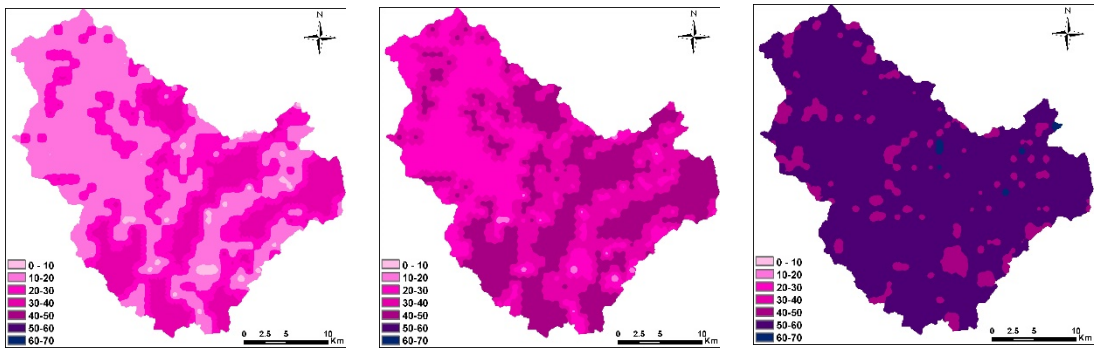


Figure S15. The final topographic index after pre-treatments.



a **b**

Figure S16. Aeration zone of Misai basin (a: before land use correction; b: after land use correction).



a b c

Figure S17. a: Field capacity; b: Wilting coefficient; c: Full storage coefficient.

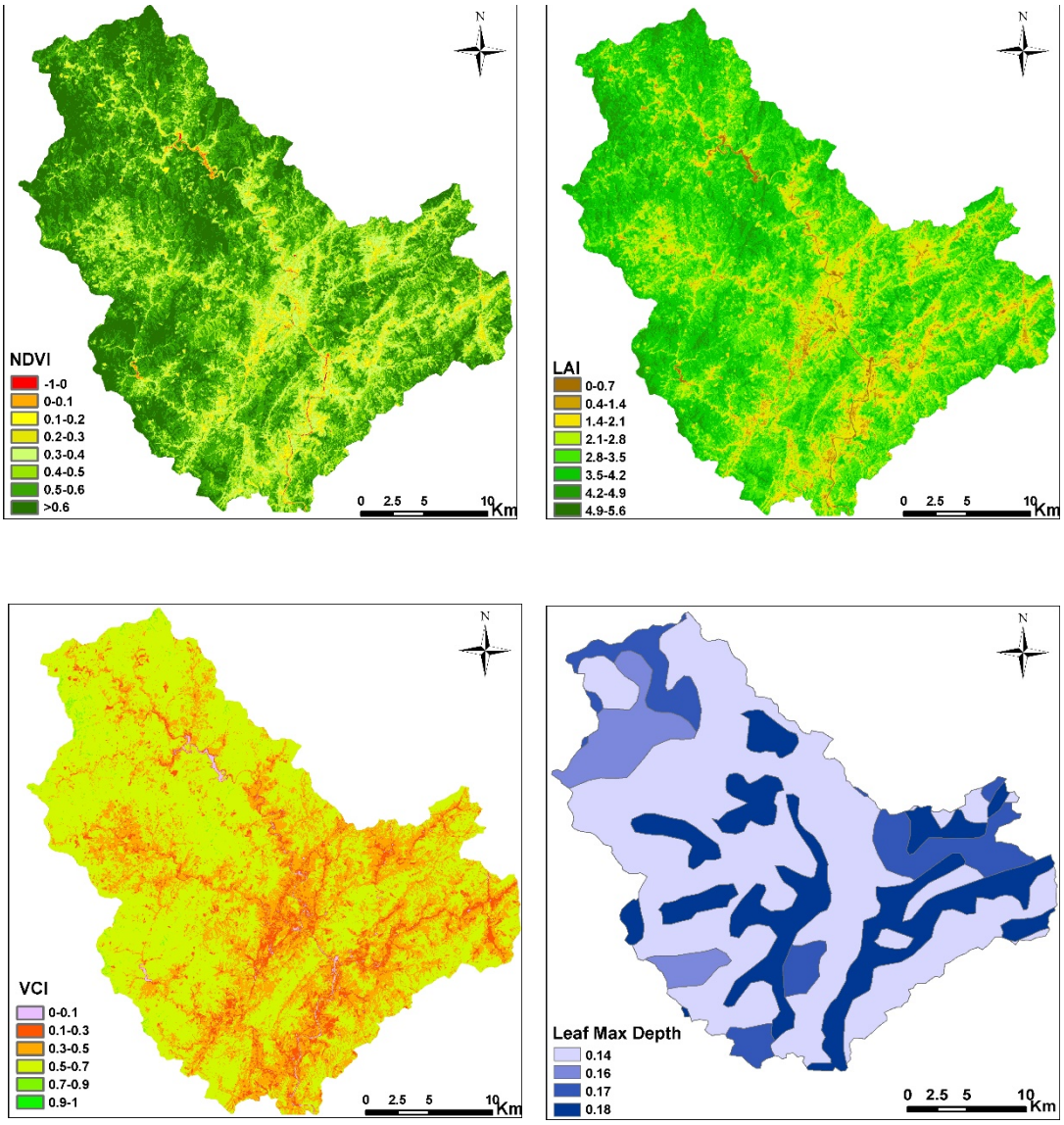


Figure S18. Vegetation information of Misai basin.

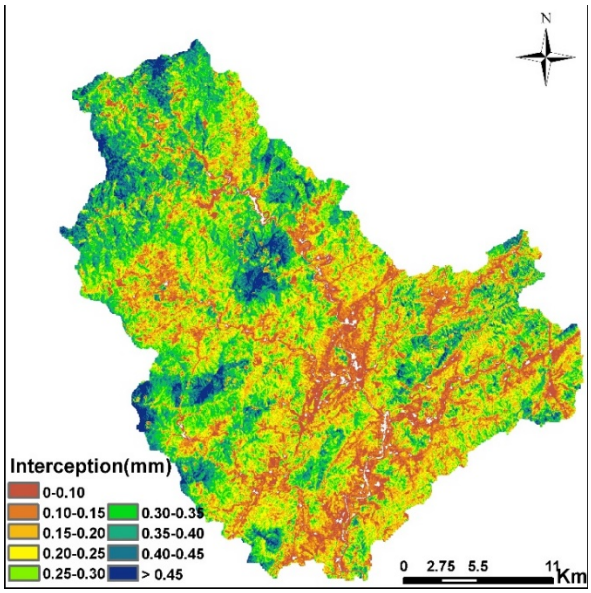


Figure S19. Canopy interception of vegetation in Misai basin.

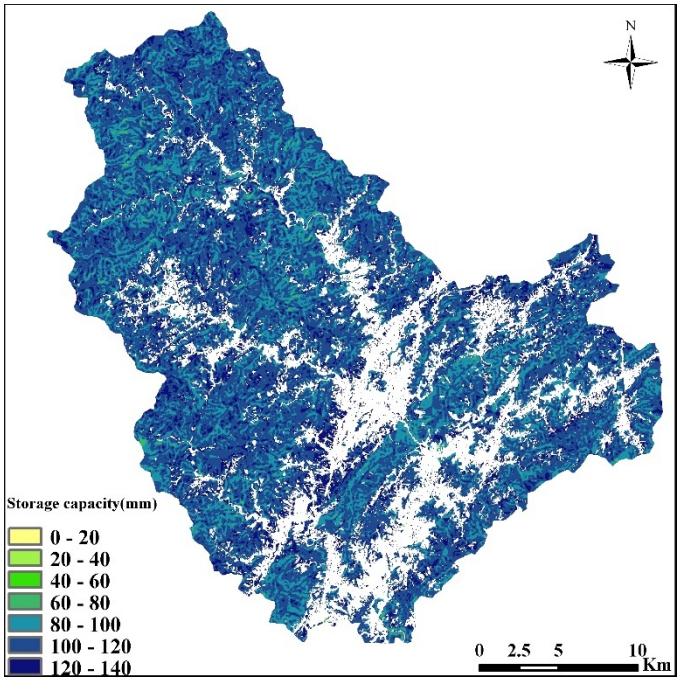


Figure S20. Spatial distribution of water storage capacity in vegetation region.

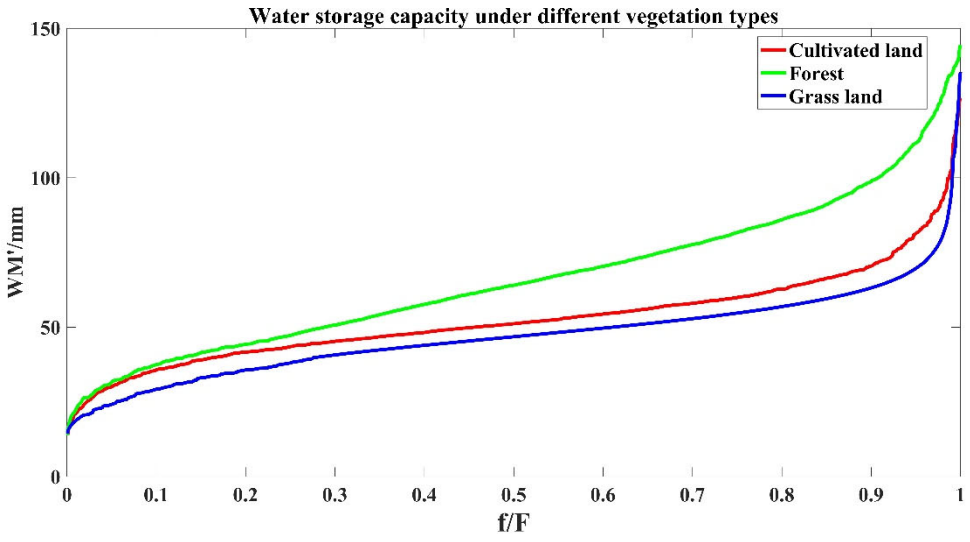


Figure S21. Water storage capacity curve under different land use in vegetation region.

Tables:

Table S1. The 17 typical flood processes (1982-1988).

Order	Start time	End time	Rainfall (mm)	Peak flow(m^3/s)
1	1982/4/2 11:00	1982/4/4 17:00	74.80	342.93
2	1982/7/17 11:00	1982/7/19 17:00	80.78	384.11
3	1982/8/5 8:00	1982/8/6 15:00	34.25	100.95
4	1983/5/29 8:00	1983/5/30 13:00	221.69	1820.04
5	1983/6/14 8:00	1983/6/16 4:00	96.90	862.09
6	1983/10/6 8:00	1983/10/7 19:00	71.70	504.99
7	1984/4/2 17:00	1984/4/4 11:00	142.55	787.04
8	1984/6/7 8:00	1984/6/9 3:00	74.22	288.91
9	1985/5/5 8:00	1985/5/7 12:00	106.15	708.00
10	1985/7/3 8:00	1985/7/5 12:00	80.12	412.00
11	1986/5/19 8:00	1986/5/22 4:00	159.20	1180.00
12	1986/7/4 8:00	1986/7/7 0:00	78.00	266.11
13	1987/4/25 8:00	1987/4/27 0:00	75.35	274.08
14	1987/5/26 8:00	1987/5/27 20:00	65.37	201.91
15	1987/5/31 8:00	1987/6/2 4:00	66.08	214.08
16	1987/9/9 8:00	1987/9/11 4:00	66.92	110.92
17	1988/6/21 8:00	1988/6/23 5:00	140.83	1410.03

Table S2. The land use information.

Land use	1982		1983		1984		1985		1986		1987		1988	
	Area(km^2)	Ratio(%)	Area(km^2)	Ratio(%)	Area(km^2)	Ratio(%)	Area(km^2)	Ratio(%)	Area(km^2)	Ratio(%)	Area(km^2)	Ratio(%)	Area(km^2)	Ratio(%)
Cultivated land	217.06	26.49	177.09	21.61	223.14	27.22	228.78	27.92	131.73	16.07	162.19	19.78	182.99	22.32
Forest land	586.19	71.54	599.71	73.17	551.53	67.29	551.47	67.30	639.70	78.05	588.27	71.76	629.73	76.81
Grass land	4.08	0.50	34.78	4.24	29.77	3.63	30.21	3.69	38.23	4.66	58.07	7.08	0.34	0.04
Water	1.56	0.19	1.59	0.19	1.57	0.19	1.53	0.19	1.56	0.19	1.60	0.20	1.59	0.19
Urban land	10.49	1.28	6.49	0.79	13.63	1.66	7.38	0.90	8.43	1.03	9.62	1.17	5.21	0.64

Table S3. The parameters of WSCC.

Module	Parameters	Physical meaning	Daily Value	Hourly Value
Evapotranspiration	WUM	Averaged storage capacity of the upper layer soil	21	35
	WLM	Averaged storage capacity of the lower layer soil	87	100
	WDM	Averaged storage capacity of the deep layer soil	37	55
	K	Evaporation coefficient	0.85	0.85
	C	Evaporation coefficient of the deep layer	0.10	0.09
Runoff generation	B	Exponential of the distribution to tension water capacity	0.50	0.42
	IMP	Percentage of impervious	0.10	0.10
runoff sources partition	SM	Free water storage capacity	17.82	100.00
	EX	Exponent of the free water capacity curve	1.00	1.02
	KG	Outflow coefficient of the groundwater	0.52	0.45
	KSS	Outflow coefficient of the interflow	0.11	0.21
Runoff routing	KKG	Depletion coefficient of the groundwater storage	0.98	0.99
	KKSS	Depletion coefficient of the interflow storage	0.88	0.99

Table S4. Grade table of model performance.

	Very good	Good	satisfactory	Unsatisfactory
NSE	>0.80	0.60≤NSE ≤0.80	0.50<NSE <0.60	≤0.50
BE	≤3.0	3. 0<BE <10.0	10.0≤BE ≤15	>15.0

Table S5. The model performance on daily scale (V, G, S and US represent very good, good, satisfactory and unsatisfactory respectively).

		NSE(L-WSCC)	NSE(WSCC)	BE(L-WSCC)	BE(WSCC)	KGE(L-WSCC)	KGE(WSCC)
Calibration	1982	0.93(VG)	0.86(VG)	2.4(VG)	-8.64(G)	0.82	0.88
	1983	0.93(VG)	0.86(VG)	-1.91(VG)	0.06(G)	0.95	0.86
	1984	0.84(VG)	0.77（G）	-6.21（G）	6.27(G)	0.87	0.7
	1985	0.81(VG)	0.84(VG)	-6.22（G）	14.2(S)	0.87	0.83
	Mean	0.88	0.83	4.19	7.29	0.88	0.82
Validation	1986	0.89(VG)	0.88(VG)	7.98(G)	14.47(S)	0.9	0.79
	1987	0.8(VG)	0.81(VG)	-1.69(VG)	7.85(G)	0.89	0.83
	1988	0.83(VG)	0.84(VG)	-3.99(G)	-27.38(US)	0.77	0.67
	Mean	0.84	0.84	4.55	16.57	0.85	0.76

Note: The mean of the BE is obtained after the absolute value.

Table S6. The model performance on hourly scale.

	Order	NSE		KGE		BE		PE		ARPT	
		WSCC	L-WSCC	WSCC	L-WSCC	WSCC	L-WSCC	WSCC	L-WSCC	WSCC	L-WSCC
Calibration	1	0.78	0.92	0.79	0.82	-5.26	-11.71	23.50	-3.31	-1	0
	2	0.68	0.73	0.54	0.75	-2.77	-11.87	-52.39	-10.65	1	1
	3	0.54	0.79	0.63	0.78	6.57	1.52	19.46	12.96	0	0
	4	0.98	0.88	0.92	0.78	1.35	4.58	-11.00	16.63	-1	0
	5	0.89	0.78	0.67	0.88	-15.52	6.40	-25.60	32.81	1	1
	6	0.67	0.65	0.70	0.64	-25.00	-29.06	-9.38	-1.97	0	1
	7	0.84	0.67	0.65	0.60	-20.62	-24.42	-17.36	-19.28	0	0
	8	0.87	0.76	0.83	0.88	-8.28	-3.29	-12.18	0.10	0	-1
	9	0.90	0.87	0.70	0.82	27.35	13.85	5.98	-3.06	0	0
	10	0.82	0.93	0.65	0.91	-16.87	-8.33	-38.80	-0.28	3	0
	Mean	0.79	0.80	0.71	0.79	12.96	11.50	21.56	10.10	-	-
Validation	11	0.95	0.94	0.80	0.84	18.54	12.41	12.37	14.67	1	1
	12	0.72	0.82	0.55	0.73	-10.87	-24.32	-47.42	-1.13	1	0
	13	0.81	0.77	0.66	0.67	15.77	-0.69	25.51	29.02	1	1
	14	0.74	0.86	0.66	0.79	2.40	-3.02	29.27	27.25	0	0
	15	0.50	0.87	0.50	0.75	10.99	6.74	28.45	18.94	-3	0
	16	0.70	0.88	0.79	0.91	3.79	-7.27	7.39	11.72	-1	-1
	17	0.88	0.83	0.74	0.79	-15.13	-17.85	-2.10	12.80	0	0
	Mean	0.76	0.85	0.67	0.78	11.07	10.33	21.79	16.50	-	-

Note: The mean of the BE/PE is obtained after the absolute value.

Table S7. The evaluation results under different types of floods (all calibration and validation period).

Flood types	NSE		KGE		BE		PE	
	WSCC	L-WSCC	WSCC	L-WSCC	WSCC	L-WSCC	WSCC	L-WSCC
B	0.98	0.88	0.92	0.78	1.35	4.58	-11.00	16.63
	0.95	0.94	0.8	0.84	18.54	12.41	12.37	14.67
	0.88	0.83	0.74	0.79	-15.13	-17.85	-2.10	12.80
M	0.89	0.78	0.67	0.88	-15.52	6.40	-25.60	32.81
	0.67	0.65	0.7	0.64	-25.00	-29.06	-9.38	-1.97
	0.84	0.67	0.65	0.6	-20.62	-24.42	-17.36	-19.28
	0.90	0.87	0.7	0.82	27.35	13.85	5.98	-3.06
S	0.78	0.92	0.79	0.82	-5.26	-11.71	23.50	-3.31
	0.68	0.73	0.54	0.75	-2.77	-11.87	-52.39	-10.65
	0.54	0.79	0.63	0.78	6.57	1.52	19.46	12.96
	0.87	0.76	0.83	0.88	-8.28	-3.29	-12.18	0.10
	0.82	0.93	0.65	0.91	-16.87	-8.33	-38.80	-0.28
	0.72	0.82	0.55	0.73	-10.87	-24.32	-47.42	-1.13
	0.81	0.77	0.66	0.67	15.77	-0.69	25.51	29.02
	0.74	0.86	0.66	0.79	2.40	-3.02	29.27	27.25
	0.50	0.87	0.5	0.75	10.99	6.74	28.45	18.94
	0.70	0.88	0.79	0.91	3.79	-7.27	7.39	11.72

Text S1: the description of WSCC

The formula of WSCC is as follows:

$$\frac{f}{F} = 1 - \left(1 - \frac{W'_m}{W'_{mm}}\right)^B \quad \text{(S1)}$$

Where, W'_m is the storage capacity of a point in the basin;f is the fraction of the basin area for which the storage capacity is less thanW'_m; F is the whole basin area; $W'_{mm}$is the maximum value of W'_m; B is the shape parameter of the storage capacity distribution.

The parameters of the storage capacity curve include the tension water capacity WM of the aeration zone, the shape factor B, and the impervious area ratio IM, which：

$$WM = \int_0^1 W'_m d(\frac{f}{F}) = \frac{W'_{mm}}{1+B} \quad \text{(S2)}$$

As shown in Fig. S10 (a), there are two basic initial assumptions for the runoff generation process: (1) the initial soil water content of the basin is W_0, and the maximum field storage capacity is A; (2) a proportion of α_0 over the basin is in the stored-full state, and the rainfall that falls on this area directly produces runoff, on the area of 1-α_0 not completely. Hence, the initial state of the basin is:

$$A = W'_{mm} * \left[1 - \left(1 - \frac{W_0}{WM}\right)^{\frac{1}{1+B}}\right] \text{(S3)}$$

If rainfall is P, evapotranspiration is E, when evapotranspiration exceeds rainfall (P − E < 0), runoff does not generate; when rainfall exceeds evapotranspiration (P − E > 0):

If P − E + A < W'_{mm}, local runoff generation, the soil water storage is the LOSS part in Fig. S10 (a) and the runoff R can be obtained by the following equation:

$$R = P - E - WM + W_0 + WM * (1 - (P - E + A)/W'_{mm})^{1+B} \text{ (S4)}$$

Otherwise:

$$R = P - E - (WM - W_0) \quad (S5)$$

The runoff yield can be calculated by equation (S4) and (S5). We can obtain the rainfall-runoff relationship as Fig. S10 (b). We can see that runoff yield only are decided by net rainfall P-E and soil moisture W_0.

Text S2. The Water Storage Capacity Calculation in Vegetation region

The vegetation region mainly includes three land use types: grassland, forest land and cultivated land. Due to the different locations of plant interception layer and deep soil, their water absorption characteristics and water storage capacity are also different. In order to improve the calculation accuracy, their water storage capacity should be calculated respectively: the vegetation interception and the extremely thin topsoil layer exposed to the air (mainly litter) are called the upper layer, and the soil layer where the vegetation roots are located is called the lower layer. In the conceptual runoff generation model, the vegetation storage process is often ignored, but for regions with good vegetation, the vegetation interception of the upper soil layer and the vegetation root water storage of the lower soil, as well as the vegetation root area, vegetation coverage and vegetation density are closely related to the distribution of water storage capacity. In the vegetation region, the changes of water storage capacity caused by these vegetation characteristics must be considered:

$$W_i = W_{u,i} + W_{l,i} \quad (S6)$$

Where W_i the storage capacity of the i-th grid is located in the vegetation region; $W_{u,i}$ is the upper layer storage capacity of the i-th grid and $W_{l,i}$ is the lower layer storage capacity of the i-th grid.

The maximum interception of vegetation is closely related to vegetation characteristics, such as leaf water holding capacity, coverage and leaf area index (LAI). According to the water storage capacity of this part, the remote sensing information model of basin interception rainfall is used to calculate the maximum interception water of vegetation:

$$I_{1,i} = f_i * LAI_i * h_i \quad (S7)$$

Where, $I_{1,i}$ is the maximum interception of the i-th grid in the vegetation region, mm; f_i is the vegetation coverage of the i-th grid; LAI_i is the vegetation leaf area index of the i-th grid, h_i is the maximum water holding capacity of vegetation leaves in the i-th grid. f_i, LAI_i and h_i can be determined by remote sensing data, and their physical meaning is clear.

A large number of experiments show a certain relationship between the water holding capacity of the litter layer and rainfall intensity and stable interception rate. The maximum water holding capacity of the litter layer is the maximum water storage capacity of the layer or by related research:

$$l_i = 0.686 * P_{max,\ i}^{\ 0.591} \quad (S8)$$

Where, l_i is the water storage capacity of litter; $P_{max,\ i}$ is the critical rainfall for the i-th grid.

The water storage capacity of soil layer is related to the ability of soil to retain water. The effective water storage capacity of soil layer is:

$$I_{2,i} = \mathrm{N}_i * \frac{\mathrm{D}_i - \mathrm{H}_i}{100} \text{ (S9)}$$

Where, $I_{2,i}$ is the water storage capacity of the i-th grid soil layer in the vegetation region(mm), N_i is non-capillary porosity (%), obtained by soil type;D_i is unsaturated zone thickness(mm), H_i is the depth of litter, generally is 50mm, obtained by vegetation type.

Hence, the calculation of water storage capacity for vegetation region is as follows:

$$\mathrm{W}_i = \mathrm{W}_{u,i} + \mathrm{W}_{l,i} = \mathrm{I}_{1,i} + l_i + I_{2,i} = \mathrm{f}_i * \mathrm{LAI}_i * \mathrm{h}_i + l_i + \mathrm{N}_i * \frac{\mathrm{D}_i - \mathrm{H}_i}{100} \text{(S10)}$$

Text S3. Data preprocessing

S3.1 Topographic index

According to the DEM data, the basin topographic index can be solved, but there must be a scale problem. Therefore, in the actual solution process, it is necessary to reduce the impact of this scale as much as possible to lay a foundation for solving the spatial distribution of water storage capacity more accurately. The topographic index at each scale is shown in Fig. S11, the corresponding frequency distribution histogram is shown in Fig. S12, and the probability density is shown in Fig. S13. From Fig. S11, S12 and S13, we can see the scale problem of the topographic index. The probability density of the topographic index after downscaling is shown in Fig. S14. The final topographic index of the basin is shown in Fig. S15.

S3.2 Thickness of aeration zone

It is an effective method to calculate the thickness of the aeration zone through the topographic index. The base flow at the outlet of the basin during drought and saturation is obtained by dividing the base flow through the measured flow of the basin and the digital filtering method, which are Q_b =2.0 m^3/s and Q_0 =5.0m^3/according to the calibration, the maximum water storage depth Szm=0.18m in the unsaturated region of Misai basin is 0.18m, so the average saturated groundwater surface depth $\overline{D}$ at the initial time of the basin is 0.12474mm. The thickness of the aeration zone obtained is shown in Fig. S16 (a).

The method of calculating the thickness of the aeration zone through the topographic index proposed above is not related to the land use, that is, the thickness of the aeration zone of the building is originally 0, but there is not through the topographic index. The thickness of the aeration zone corrected by land use is shown in Fig. S16 (b).

S3.3 Soil data processing

Using the soil data provided by the Food and Agriculture Organization of the United Nations (FAO) (Fig. S3 a) ,select the relevant soil attribute distribution data of the basin, including the percentage content of soil clay, sand and organic matter in the study area. Based on the SPAW software, the field capacity and wilting coefficient in each grid are calculated (Fig. S17).

S3.4 Vegetation data processing

According to the vegetation type map of Misai basin (Fig. S3 b), the NDVI, leaf area index LAI, vegetation coverage VCI and maximum leaf water holding capacity LMD can be solved, as shown in Fig. S18. Based on the above data and the thickness of the aeration zone, the canopy

interception in the vegetation regions is obtained, as shown in Fig. S19, the spatial distribution of the water storage capacity is shown in Fig. S20, and the water storage capacity curve under different land uses in the vegetation region is shown in Fig. S21.

Text S4. Statistical Criteria and Model Evaluation Performance

The performance of the WSCC and L-WSCC can be computed using statistical indices and graphical comparisons. For the daily simulated discharge, the Nash–Sutcliffe efficiency (NSE), percent bias error (BE), and Kling–Gupta Efficiency (KGE) were adopted to evaluate the daily model performance. For the hourly simulated discharge, the difference of appearance time of peak (ARPT) and flood peak error (PE) also are added to evaluate the performance. In order to study the degree of influence of different factors on runoff yield, the change of mean absolute error (ΔMAE) is also used.

$$\mathrm{NSE}=1-\frac{\sum(Q_o-Q_s)^2}{\sum(Q_o-\overline{Q_o})^2} \quad \text{(S11)}$$

$$\mathrm{BE}=\frac{\sum Q_s-\sum Q_o}{\sum Q_o}\times 100\% \quad \text{(S12)}$$

$$\mathrm{PE}=\frac{\mathrm{MAX}(Q_s)-\mathrm{MAX}(Q_o)}{\mathrm{MAX}(Q_o)}\times 100\% \quad \text{(S13)}$$

$$\mathrm{ARPT}=T_1-T_2 \quad \text{(S14)}$$

$$\mathrm{KGE}=1-\sqrt{(r-1)^2+(\alpha-1)^2+(\beta-1)^2} \quad \text{(S15)}$$

$$\alpha=\frac{\sigma_s}{\sigma_o} \quad \beta=\frac{\overline{Q_s}}{\overline{Q_o}} \quad \text{(S16)}$$

$$\Delta\text{MAE}=\frac{\sum|\mathrm{R_{adj}}-\mathrm{R_{L\text{-}WSCC}}|}{m}\times 100\%\ \text{(S17)}$$

Where Q_o is the observed discharge, Q_s is the simulated discharge, $\overline{Q_o}$ is the mean of observed discharge, $\overline{Q_s}$ is the mean of simulated discharge, $\mathrm{MAX}(\mathrm{Q_s})$ is the simulated peak discharge, $\mathrm{MAX}(\mathrm{Q_o})$ observed peak discharge,T_1 is the simulated flood peak appearance time, T_2 is the observed flood peak appearance time, σ_s is the standard deviation of simulated discharge, σ_o is the standard deviation of observed discharge, r is the correlation coefficient, $\mathrm{R_{adj}}$ is the simulated runoff yield in different scenarios, $\mathrm{R_{L\text{-}WSCC}}$ is the simulated runoff yield of L-WSCC, m is the number of data.