Journal of Hydrology 562 (2018) 1–16

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Research papers

Influence of rainfall data scarcity on non-point source pollution prediction: Implications for physically based models

Lei Chen[a], Jiajia Xu[a], Guobo Wang[b], Hongbin Liu[c], Limei Zhai[c], Shuang Li[a], Cheng Sun[a], Zhenyao Shen[a,*]

[a] State Key Laboratory of Water Environment Simulation, School of Environment, Beijing Normal University, Beijing 100875, PR China
[b] Waterborne Transport Research Institute, Ministry of Transport of the People's Republic of China, Beijing 100088, PR China
[c] Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Key Laboratory of Non-Point Source Pollution Control, Ministry of Agriculture, Beijing 100081, PR China

ARTICLE INFO

This manuscript was handled by G. Syme, Editor-in-Chief, with the assistance of Hazi Mohammad Azamathulla, Associate Editor

Keywords:
Information entropy
Non-point source pollution
Rainfall data
Scarcity
Soil and Water Assessment Tool
Three Gorges Reservoir Region

ABSTRACT

Hydrological and non-point source pollution (H/NPS) predictions in ungagged basins have become the key problem for watershed studies, especially for those large-scale catchments. However, few studies have explored the comprehensive impacts of rainfall data scarcity on H/NPS predictions. This study focused on: 1) the effects of rainfall spatial scarcity (by removing 11%–67% of stations based on their locations) on the H/NPS results; and 2) the impacts of rainfall temporal scarcity (10%–60% data scarcity in time series); and 3) the development of a new evaluation method that incorporates information entropy. A case study was undertaken using the Soil and Water Assessment Tool (SWAT) in a typical watershed in China. The results of this study highlighted the importance of critical-site rainfall stations that often showed greater influences and cross-tributary impacts on the H/NPS simulations. Higher missing rates above a certain threshold as well as missing locations during the wet periods resulted in poorer simulation results. Compared to traditional indicators, information entropy could serve as a good substitute because it reflects the distribution of spatial variability and the development of temporal heterogeneity. This paper reports important implications for the application of Distributed Hydrological Models and Semi-distributed Hydrological Models, as well as for the optimal design of rainfall gauges among large basins.

1. Introduction

Non-point source (NPS) pollution has been a key threat to water quality for decades, and hydrological and non-point source (H/NPS) models are the main tools used to quantify NPS pollution (Andréassian et al., 2001; Bera and Borah, 2003; Chaubey et al., 1999; Drecht et al., 2003). Typically, H/NPS models can be divided into Lumped Hydrologic Models (LHMs), Distributed Hydrological Models (DHMs), and Semi-distributed Hydrological Model (SDHMs). The LHMs regard the watershed as a whole object and cannot reflect the spatial heterogeneity of the actual process inside the watershed (Hrachowitz and Clark, 2017). Conversely, DHMs/SDHMs divide watersheds into smaller spatial units, while pollutants are calculated from each separate unit and are then summed at the watershed outlet. As one special kind of DHMs, the SDHMs divided the entire basin into sub-watersheds first and then into a number of hydrological response units (HRUs) or other computational units depending on slope, soil type and land use instead of rectangular grids (with uniform size) (Hrachowitz et al., 2016; Viviroli et al., 2009). In this sense, spatial variations in climate, underlying surfaces and related hydrological elements could be considered, and spatial data with higher accuracy that are typically derived via remote sensing (RS) and geographic information system (GIS) technologies are required (Bieger et al., 2014; Wang et al., 2016). The commonly used DHMs/SDHMs include the Soil and Water Assessment Tool (SWAT) model, the Institute of Hydrology Distributed Model (IHDM) and TOPMODEL. Among these models, the SWAT model has become one of the most widely used tools in describing temporal and spatial variations in H/NPS cycles, especially for large-scale watersheds due to their greater heterogeneities (Amatya et al., 2011).

Rainfall data are regarded as the most important inputs for DHMs/SDHMs because they act as the driving force of runoff generation and pollutant transportation (Lobligeois et al., 2014; Kashani et al., 2016; Sun et al., 2017). Typically, rainfall data could be obtained using both rainfall station and radar product (Kashani et al., 2016; Pereiracardenal et al., 2011). The application of the radar product has become more widespread as radar technique can reflect the spatial and temporal

* Corresponding author.
E-mail addresses: chenlei1982bnu@bnu.edu.cn, zyshen@bnu.edu.cn (Z. Shen).

https://doi.org/10.1016/j.jhydrol.2018.04.044
Received 4 February 2018; Received in revised form 15 April 2018; Accepted 17 April 2018
Available online 30 April 2018
0022-1694/

rainfall variability, especially for those remote regions (Biggs and Atkinson, 2015). However, the radar rainfall product is sometime questioned as its resolutions are generally based on quantitative rainfall estimates (QPEs) from weather radar networks (Harrison et al., 2009). Several major disadvantages of QPEs, such as low spatial resolution, low level forecast for torrential rain and rough prediction of spatial-temporal structure of heavy rainfall, make the radar data cannot meet the requirements of H/NPS models (Cluckie et al., 2004). Thus, there is a growing need for combined (synergic) use of radar measurements along with the rainfall stations, and the rain gauge is still the fundamental source used by DHMs/SDHMs to determine rainfall variability.

Previous studies have indicated that rainfall has irregular changes due to the varying natural conditions and shows strong spatial-temporal heterogeneities among large-scale watersheds, which would impact simulated results of the DHMs/SDHMs (Bardossy and Plate, 1992; Shen et al., 2012). Long time series of rainfall data is often required for comprehensive H/NPS evaluations by covering high flow, normal flow and low flow periods to the greatest extent (Kuangyao et al., 2000; Troin et al., 2012). However, automatic weather stations are susceptible because certain time series may be lost in part due to direct and indirect damage caused by lightning electromagnetic pulses. Electromagnetic interference and human operational errors are also destructive factors for automatic stations and can result in random and continuous absences of rainfall data. As a result, damage to automatic stations and other equipment errors due to various human and natural factors can cause losses in rainfall data series (Shen et al., 2015a; Wambura et al., 2017). On the other hand, previous study has indicated spatial rainfall variability is an important source of NPS simulation uncertainty; thus, users must focus on the optimal design of rainfall stations (Shen et al., 2012). However, accurate descriptions of spatial rainfall variability cannot be obtained due to the paucity of existing observation sites, which is caused by economic and technique conditions (i.e., terrain). The spatial and temporal resolutions of rainfall data are constrained and interrelated and will affect the quantification of H/NPS (Meusburger et al., 2012; Michaelides et al., 2009). Thus, rainfall data scarcity indeed exists and has become the key barrier to H/NPS prediction.

H/NPS predictions in ungauged basins (PUB) have become a hot topic for hydrological researchers because data scarcity is still a worldwide problem. The International Association of Hydrological Sciences (IAHS) began the PUB programme as its first key research programme in the 21st century, which is called as 'PUB, 2003–2012: Shaping an exciting future for the hydrological sciences.' (Hrachowitz et al., 2013). For those DHMs/SDHMs, scarce scenarios of rainfall data could be divided into two types: spatial series and time series. Satellite remote sensing play an important role in the PUB (Lakshmi, 2004), but it cannot solve the problem of high spatial resolution and frequent time duplication (Lakshmi, 2016). Studies have also focused on the effect of different spatial interpolation methods and their combinations on rainfall estimation during different rainfall periods (Ali et al., 2005; Cheng et al., 2017; Swain and Patra, 2017). Although these studies have shown that spatial rainfall variability could be obtained by interpolating rainfall data of each station by the methods such as Thiessen polygons, average method and centroid method (Cho et al., 2009), there are few studies focused on their performances during data scarcity scenarios specifically. Weather generators have been developed for DHMs/SDHMs to cope with data scarcity in time series, but the impacts of data scarcity scenarios on their interpolation abilities are not clear yet (Chen et al., 2017a; Ruan et al., 2016). In that regard, quantifying the impacts of the temporal and spatial scarcity of rainfall data is indeed crucial because those data provide the basic inputs for H/NPS models.

Therefore, we focused on the influences of rainfall data scarcity on total phosphorus (TP) prediction in a NPS-dominant catchment by running and comparing different rainfall scarcity scenarios. The following tasks have been performed by: 1) quantifying the effects of rainfall spatial scarcity (station number and location) on H/TP simulations, and 2) exploring the impacts of rainfall temporal scarcity (data scarcity in time series) on H/TP predictions, and 3) developing the information entropy as a new method to replace traditional indicators for data scarcity evaluation. The study is carried out for Daning watershed, China.

2. Methods and materials

2.1. Study area description and data collection

2.1.1. Study area description

The Daning River watershed is located in Wushan county and Wuxi county, the Municipality of Chongqing, China. As an important tributary in the Three Gorges Reservoir area, the drainage area of the Daning River watershed is approximately 2422 km^2 and consists of four major tributaries, including the Xixi River, the Houxi River, the Boyang River and the Dongxi River. The mean rainfall is 1030–1950 mm per year and the annual temperature of the watershed is approximately 18.4 °C. The elevation of the Daning River watershed ranges from 200 m to 2605 m, while the land use types in this area are primarily comprised of forest lands (61.85%), croplands (24.90%) and grasslands (12.48%). This area also consists of seven major soil types: yellow soils (46.05%), yellow-brown soils (25.79%), limestone soils (18.19%), brown soils (6.28%), purple soils (1.98%), paddy soil (1.52%) and alluvial soils (less than 0.2%). The location of the watershed in the Three Gorges Reservoir area and other information are shown in Fig. 1. Based on historical record, the Daning River suffers serious NPS pollution, and phosphorus (P) is the limiting nutrient causing eutrophication in the Three Gorges Reservoir Region (Shen et al., 2012, 2015b). In this area, excessive use of phosphate fertilizer are used on slope cropland and the rainfall would cause P being carrying into the receiving body of water through rainfall erosion and runoff transport from non-specific sites in form of NPS pollution. Thus, TP was selected as a representative of NPS pollutants and water quality in this region.

2.1.2. Data collection

For a comprehensive evaluation, detailed data that are available in this region are collected and compiled as follows:

- Daily rainfall data at nine rainfall stations inside the watershed and four stations at sites approximately 30 km outside the watershed boundary from 1998 to 2008 were obtained from the Meteorological Bureau of Wuxi County and China National Meteorological Administration.
- Other meteorological data from 2000 to 2008, such as daily maximum and minimum air temperature, relative humidity, wind speed and sunlight radiation, were collected at the Meteorological Bureau of Wuxi County.
- The Digital Elevation Model (DEM), which had a resolution of 1:250,000, was digitized from raw data provided by the National Fundamental Geographic Information Center of China.
- The land use map, which had a resolution of 1:100,000, was obtained from the Resources and Environment science data Center of the Chinese Sciences Academy.
- The soil type map, which had a resolution of 1:50,000, was obtained from the Agricultural Science Committee of Wuxi city. The soil physical properties, including soil density, saturated and unsaturated soil hydraulic conductivity, and field capacity, were acquired from the Institute of Soil Science in Nanjing. The soil chemical properties were obtained from the Soil Database of China.
- The crop management measures were obtained from field investigations with local farmers.
- The measured daily flow data and monthly water quality data are mainly obtained from the Yangtze River Basin Water Conservancy Commission and the Wuxi County Environmental Protection

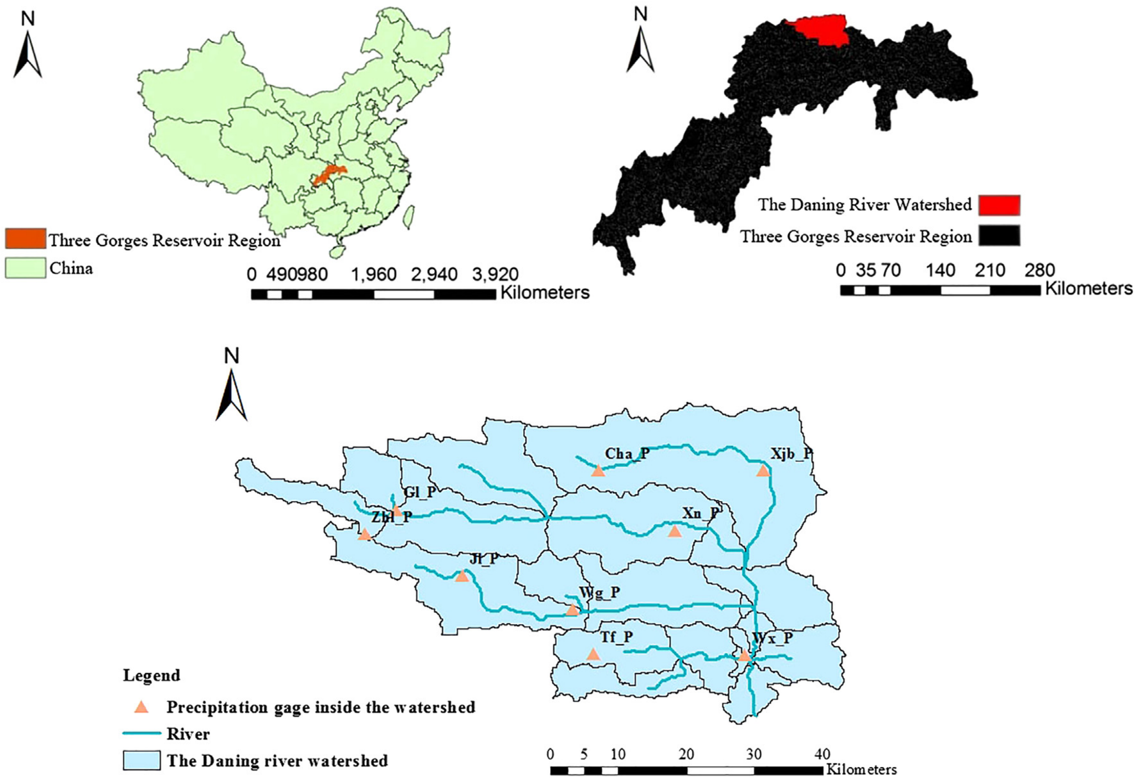

Fig. 1. Location of the Wuxi section in the Daning River watershed.

Bureau.

2.2. Setup of the SWAT model

2.2.1. Model description

As a physical-based and semi-distributed model, the SWAT is used for the H/NPS prediction in this study. The SWAT represents spatial heterogeneities of land covers, soil types and management practices and is therefore suitable for large-scale and complex watershed studies (Douglas-Mankin et al., 2010). The Soil Conservation Service (SCS) method (1972) and Green & Ampt infiltration methods (1911) are the two alternative methods for surface runoff estimation, while the Penman-Monteith, Priestley-Taylor and Hargreaves methods are used for the evapotranspiration estimation. The simulation of subsurface flow reference from a kinematic storage method, and the routing channel water is simulated by the Muskingum method. The Modified Universal Soil Loss Equation (MUSLE) is incorporated to simulate soil erosion and estimate sediment yield.

The SWAT can monitor organic P, inorganic P and soluble P and simulates the P cycle in the water stream. Specifically, the solubility of P is limited in most environments and this property of P makes it accumulate near the soil surface and migrate easily with surface runoff (Heathwaite and Dils, 2000; Smil, 2000; Sharpley et al., 2001). Moreover, the SWAT can simulate six P pools in the soil, including three inorganic and three organic forms, such as insoluble forms of mineral P and P in the soil solution etc.

$$P_{surf} = 0.001 \times C_{orgP} \cdot \frac{Q_{sed}}{A_{hru}} \cdot \varepsilon_{P:sed} \tag{1}$$

where P_{surf} is the amount of organic P in surface runoff, and C_{orgP} is the concentration of organic P in the top 10 mm of soil, and $\varepsilon_{P:sed}$ is the P enrichment ratio, and A_{hru} is the area of the HRU, and Q_{sed} is the sediment yield on a given day.

$$Q_{sed} = 11.8(Q_{surf} \cdot q_{peak} \cdot A_{hru})^{0.56} \cdot K_{ulse} \cdot C_{ulse} \cdot P_{ulse} \cdot L_{ulse} \cdot F_{CFRG} \tag{2}$$

where Q_{sed} is the sediment yield on a given day; Q_{surf} is the surface runoff volume; q_{peak} is the peak runoff rate; A_{hru} is the area of the HRU; K_{usle}, C_{usle}, P_{usle} and L_{usle} is the USLE soil erodibility factor, cover and management factor, topographic factor; and coarse fragment factor, respectively.

2.2.2. Rainfall module of the SWAT

To supplement incomplete meteorological data in temporal series, the weather generator is incorporated as a module of the SWAT (Hartkamp et al., 2003; Wilks, 2009). The weather generator could be used to interpolate the scarce data in temporal series, and six meteorological variables are interpolated from the known high and low values, including rainfall, highest temperature, lowest temperature, daily average relative humidity, daily radiation and daily average wind speed (Kilsby et al., 2007). The daily rainfall data are simulated using a Markov Chain - skewed distribution model (Richardson, 1981). According to the skewed distribution function of the SWAT model, the formula for calculating the rainfall in rainy days is:

$$R_{day} = \mu_{mon} + 2\sigma_{mon} \frac{\left[\left(SND_{day} - \frac{g_{mon}}{6}\right)\left(\frac{g_{mon}}{6}\right) + 1\right]^3 - 1}{g_{mon}} \tag{3}$$

where R_{day} is the one-day rainfall, μ_{mon} is the average daily rainfall in one month, σ_{mon} is the standard deviation of the average daily rainfall in one month, SND_{day} is the standard normal deviation of the calculated one-day, and g_{mon} is the skewness coefficient of the average daily rainfall in one month.

The standard normal deviation for one day is

$$SND_{day} = \cos(6.283 rnd_2)\sqrt{-2\ln(rnd_1)} \tag{4}$$

where rnd_1 and rnd_2 are random numbers between 0 and 1.0.

To capture the spatial rainfall variability, Thiessen polygons are used for the SWAT simulation in the large watershed and the SWAT also uses elevation bands to account for orographic effects on rainfall and temperature. The Thiessen polygons depend on the overall average or data near the missing data and the elevation bands are used in conjunction with other methods to calculate rainfall and simulate runoff in the mountain location (Fontaine et al., 2002; Zhang et al., 2008).

2.2.3. Model setup

Using the DEM, the Daning River watershed was divided into 22 sub-watersheds based on a threshold of the critical source area of 5000 ha. The sub-watersheds were then divided into HRUs according to the different underlying surface features, including land use, soil type and slope. The Sequential Uncertainty Fitting Version-2 (SUFI-2) of the SWAT Calibration and uncertainty programme was used for calibration and validation (Abbaspour et al., 2007). According to the data obtained, the calibration and verification in this paper are based on the monthly time step. The hydrological and water quality data from 2004 to 2008 are selected as calibration periods because of its completeness. Hydrological and water quality data from 2000 to 2003 are selected as validation periods. A one-year warm-up period (1999) was specified for both the calibration and validation periods. Traditional methods were used for model evaluation (Merz and Blöschl, 2004), including the correlation coefficient R^2 (Legates and McCabe, 1999) and the Nash-Sutcliffe coefficient (Ens) (Nash and Sutcliffe, 1970).

The correlation coefficient R^2

$$R^2 = \left[\left(\sum_{i=1}^{n}(O_i-\overline{O})(P_i-\overline{P})\right)\Big/\left(\sqrt{\sum_{i=1}^{n}(O_i-\overline{O})}\sqrt{\sum_{i=1}^{n}(P_i-\overline{P})^2}\right)\right]^2 \tag{5}$$

where O_i is the ith observed value, P_i is the ith simulated value, is the average of the observed values, n is the number of observed values (or simulated values). When the simulated values are equal to the measured values, then $R^2 = 1$. When $R^2 < 1$, the smaller the value are, the lower the degree of data fit.

The Nash-Sutcliffe coefficient (Ens)

$$Ens = 1-\sum_{i=1}^{n}(O_i-P_i)^2/\sum_{i=1}^{n}(O_i-\overline{O})^2 \tag{6}$$

when the simulated values are equal to the measured values, Ens = 1 then. When Ens < 1, the smaller the value of a reflects the lower degree of data coincidence. When Ens is less than 0, it indicates that the model performance is poor.

In sum, the R^2 values were quantified as 0.79 and 0.95 for the simulated flows and TP, respectively, and the Ens values were 0.74 and 0.93, respectively, indicating that the SWAT model performed well in this study area. For more information about this process, please see our previous studies (Chen et al., 2014; Gong et al., 2012).

2.3. Design of the rainfall data scarcity

2.3.1. Spatial data scarcity

In this study, both spatial and temporal data scarcity are designed and the details of data scarcity scenario could be found in Table 1. In the study of spatial data scarcity, the daily rainfall data from 2000 to 2008 were obtained from 9 stations, which are the Xining, Changan, Xujiaba, Gaolou, Zhongliang, Jianlou, Wangu, Tangfang and Wuxi rainfall station (Fig. 1). The spatial rainfall variation was then generated for the 22 sub-watersheds by regarding each station as a single unit and interpolating them using the Thiessen polygon method. Therefore, the baseline scenario is setup by using all complete datasets and the spatial data scarcity is then designed by reducing the number of rainfall stations and their spatial locations. For data-scarce scenarios, the number of rainfall stations is often limited so the impacts of station number on NPS-TP load prediction was compared firstly.

To make more comprehensive results, information entropy was used to evaluate the spatial model performances in addition to *Ens*. Entropy theory, which was introduced by Shannon in the 1940s, is widely used in the fields of hydrology and the environment (Bian and Feng, 2010; Khosravi et al., 2016). Shannon notes that there is redundancy in any information, and the size of the redundancy is related to the probability or uncertainty of each symbol (number, letter, or word) in the message. The information entropy is the amount of information which is the average amount of information after excluding redundancy (Ilunga, 2017). Some studies have shown that the information entropy method has good results with less computational complexity to compute the fuzzy formal concepts (Shemshadi et al., 2011; Singh et al., 2014). In this study, entropy theory is used in the following forms to solve a variety of problems in terms of H/NPS model evaluation (Chen et al., 2017b).

$$H_{spatial}(x_j) = -\sum_{i=1}^{n} P(x_{ij})logP(x_{ij}) \tag{7}$$

$$H_{temporal}(x_i) = -\sum_{j=1}^{m} P(x_{ij})logP(x_{ij}) \tag{8}$$

where $H_{spatial}(x_j)$ is the information entropy of each sub-watershed in the whole simulation period, and $H_{temporal}(x_i)$ is the information entropy of the whole watershed in monthly time step, x_{ij} is sub-watershed j for month *ith*, and $P(x_{ij})$ is the ratio of the simulated flow or pollutants data at x_{ij} to the total output of the entire basin in monthly time step, and n is the number of months during simulations, and m is the number of sub-watersheds in the whole watershed.

Then indicator of ΔH is used as means that deviation in entropy values between baseline scenario and data scarcity scenario. In this sense, the impacts of each single rainfall station on H/NPS prediction could be reflected by the information theory. Then importance of each paired stations on the model simulation results could be ranked by ΔH value during each scenario. This study employed six scenarios by gradually decreasing the total entropy of the rainfall stations (that is, station numbers) and more details about six scenarios are shown in Table 1. The scenarios included a complete scenario with 9 rainfall stations (named S0), an 8-gauge scenario that lacked the Xining station (named S1), and a 7-gauge scenario that lacked the Changan and Xujiaba stations (named S2). Furthermore, there were 5-gauge, 4-gauge and 3-gauge scenarios, wherein the number of rainfall stations was gradually reduced (named S3, S4 and S5 respectively).

By comparing the simulated values, those important stations, which showed greater influences on the H/NPS simulations, are defined as the critical-site rainfall stations. Besides, the importance of the rainfall stations in a watershed is related to the amount of information contained in the river basin (the basin area is controlled by a single rainfall station) and the locations of the rainfall stations. Specifically, upstream P would transmit to its downstream reaches and affect a series of physicochemical processes of P in the downstream channel. Thus, the impacts of scarcity data location at the upper reaches was further analysed based on a concept of influence sphere by comparing ΔH values of the simulated H/TP data of several related downstream sub-watersheds. Therefore, this paper uses tributaries as confluence units to study the actual influence of every rainfall station on a series of downstream reaches and establish the importance of each station location by quantifying their downstream ΔH values. To avoid the influence of time series scarcity on the simulation results, seven rainfall stations (Gaolou, Jianlou, Tangfang, Wangu, Wuxi, Xining and Zhongliang) that had relatively complete data were chosen to analyse the influence sphere of the each station location. These seven scenarios are named S6–S12 and more details about scenarios settings can be found in Table 1.

2.3.2. Temporal series scarcity

To reflect the actual data scarcity in the Daning River watershed, an additional four stations in addition to the above nine stations used for spatial data scarcity analysis were used to accurately analyse the actual scarcity situation. Daily rainfall data from 13 rainfall stations on the

Table 1
The experimental design of rainfall data scarcity scenario.

Category		Name	Detailed description
Baseline		S0	No scarce data scenarios (no lack of time series and also retained a complete nine stations)
Spatial data scarcity	Decreasing number of rainfall stations	S1	An 8-gauge scenario that lacked the Xining station
		S2	A 7-gauge scenario that lacked the Changan and Xujiaba stations
		S3	A 5-gauge scenario that lacked the Changan, Xujiaba, Wangu and Zhongliang stations
		S4	A 4-gauge scenario that lacked the Changan, Xujiaba, Wangu, Zhongliang and Gaolou stations
		S5	A 3-gauge scenario that lacked the Changan, Xujiaba, Wangu, Zhongliang, Gaolou and Tangfang stations
	Effect of the location of a removed rainfall stations	S6	A scenario with 6 stations that lacked the Gaolou station
		S7	A scenario with 6 stations that lacked the Jianlou station
		S8	A scenario with 6 stations that lacked the Tangfang station
		S9	A scenario with 6 stations that lacked the Wangu station
		S10	A scenario with 6 stations that lacked the Wuxi station
		S11	A scenario with 6 stations that lacked the Xining station
		S12	A scenario with 6 stations that lacked the Zhongliang station
Temporal data scarcity (in the Xining station)	Rainfall time series degradation with increasing missing period	S13	A scenario with 10% data scarcity
		S14	A scenario with 20% data scarcity
		S15	A scenario with 30% data scarcity
		S16	A scenario with 40% data scarcity
		S17	A scenario with 50% data scarcity
		S18	A scenario with 60% data scarcity
	Rainfall time series degradation with variable timing of the missing period	S19	Pattern 1 with data scarcity in the high flow year of 2000
		S20	Pattern 2 with data scarcity in the normal flow year of 2002
		S21	Pattern 3 with data scarcity in the low flow year of 2004
		S22	Pattern 4 with data scarcity in the high flow year of 2005
		S23	Pattern 5 with data scarcity in the high flow year of 2007

Daning River from 1998 to 2008 are therefore used to represent the actual data scarcity in time series. There were no daily rainfall data from 1998 to 2000 because the Wangu and Zhongliang stations were established after 2000. The Longmen, Changan and Shuangyang stations did not provide continuous rainfall data in 2003 and 2008 due to a state of disrepair. The Gaolou station has continuous time series data scarcity in 2003 and short time series scarcity in 2008. Short durations of daily rainfall were missing from the Wangu and Xujiaba stations in 2007. The Tangfang and Jianlou stations yielded scarce data over 2003. In summary, the rainfall data for the Daning River is available at random times and is for the most part discontinuous and that some short time series are missing. The actual situation of scarce rainfall data is shown in Fig. 2.

To avoid spatial scarcity mentioned above, the Xining station, which had complete data series, was used for experiment design of temporal scarcity condition. In this study, different scarce data scenarios were designed according to the actual situation of the watershed. To quantify the impacts of data scarcity, the scarce data scenarios were divided into different missing rates and different missing positions (Lopez and Seibert, 2016) and the missing data distributions with different missing rates and different missing positions are shown in Fig. 2(a). The different missing scenarios had random and continuous scarce data in time, with missing short time series. For this paper, the six missing rates ranged from 10% to 60%. The settings of six missing rates scarcity scenario named S13 to S18 are shown in Fig. 2(b) and Table 1. The minimum missing rate of 10% was used for the end of 2004, the beginning of 2005 and for most of 2007. A maximum missing rate of 60% was utilized for the beginning of 2000, the ends of 2001 and 2008, for some parts of 2004 and 2005, and for the entireties of 2002, 2003, 2006 and 2007. As the six missing rates increased, the scarce time periods for the remaining missing rates increased. In addition, Fig. 2(c) shows five different missing positions (patterns 1–5). Among them, patterns 1, 4 and 5 represent the data scarcity of 2000, 2005 and 2007 which are the high flow years, respectively. Pattern 2 represents the data scarcity of 2002 year which is the normal flow year. Pattern 3 was designed to represent the data scarcity of 2004 year which is the low flow year.

3. Results

3.1. Impacts of spatial data scarcity

3.1.1. Impacts of station number and location

The H/TP evaluations results during data spatial scarcity conditions are shown in Table 2. Specifically, the information entropy of the Gaolou, Zhongliang, Jianlou, Xining, Wuxi, Wangu, Tangfang, Xujiaba and Changan station is quantified as 0.2002, 0.0906, 0.2756, 0.3638, 0.2410, 0.0757, 0.1656, 0.3111 and 0.1909. And information entropy for S0 to S5 was 1.9145, 1.5507, 1.4125, 1.2462, 1.0460 and 0.8804 in turn, while the Ens of flow and TP were reduced from 0.7425 to 0.6023 and from 0.9299 to 0.742, respectively. This indicates that the model performance gradually deteriorated if spatial rainfall scarcity exists. In compare, the 8-gauge scenario lacked the main site of the Xining station, which reduced the total set of information by approximately 25%, indicating an obvious decrease of model performance during the 8-gauge scenario. However, the simulation results from the several scenarios that retained the rainfall station with the highest information entropy were satisfactory. The resulting effects using the S1 and S2 scenarios were identical because although the Changan and Xujiaba stations were removed during the S2 scenario, these two stations were not taken into account due to the interpolation using the Thiessen polygons in SWAT. The $\Delta H_{Temporal}$ values for the simulated H/TP are shown in Fig. 3. It could be observed that the fluctuations in $\Delta H_{temporal}$ were relatively stable during different scenarios by mainly concentrating in the average value and its vicinity. Compare to Ens values that reflected model performance at the catchment outlet, the $\Delta H_{temporal}$ could better reflect the simulation results and their responses to rainfall data scarcity. As the information entropy decreased from 1.2462 to 0.8804, the maximum and minimum values of the fluctuations increased from 0.41 to 0.86 and from 0.11 to 0.52, respectively, which reflected the gradual increase in fluctuation range of $\Delta H_{temporal}$ as the total information entropy decreased. In addition, the maximum and minimum fluctuations during S1 scenario, which did not contain large information entropy, were 0.89 and 0.33, respectively. In other words,

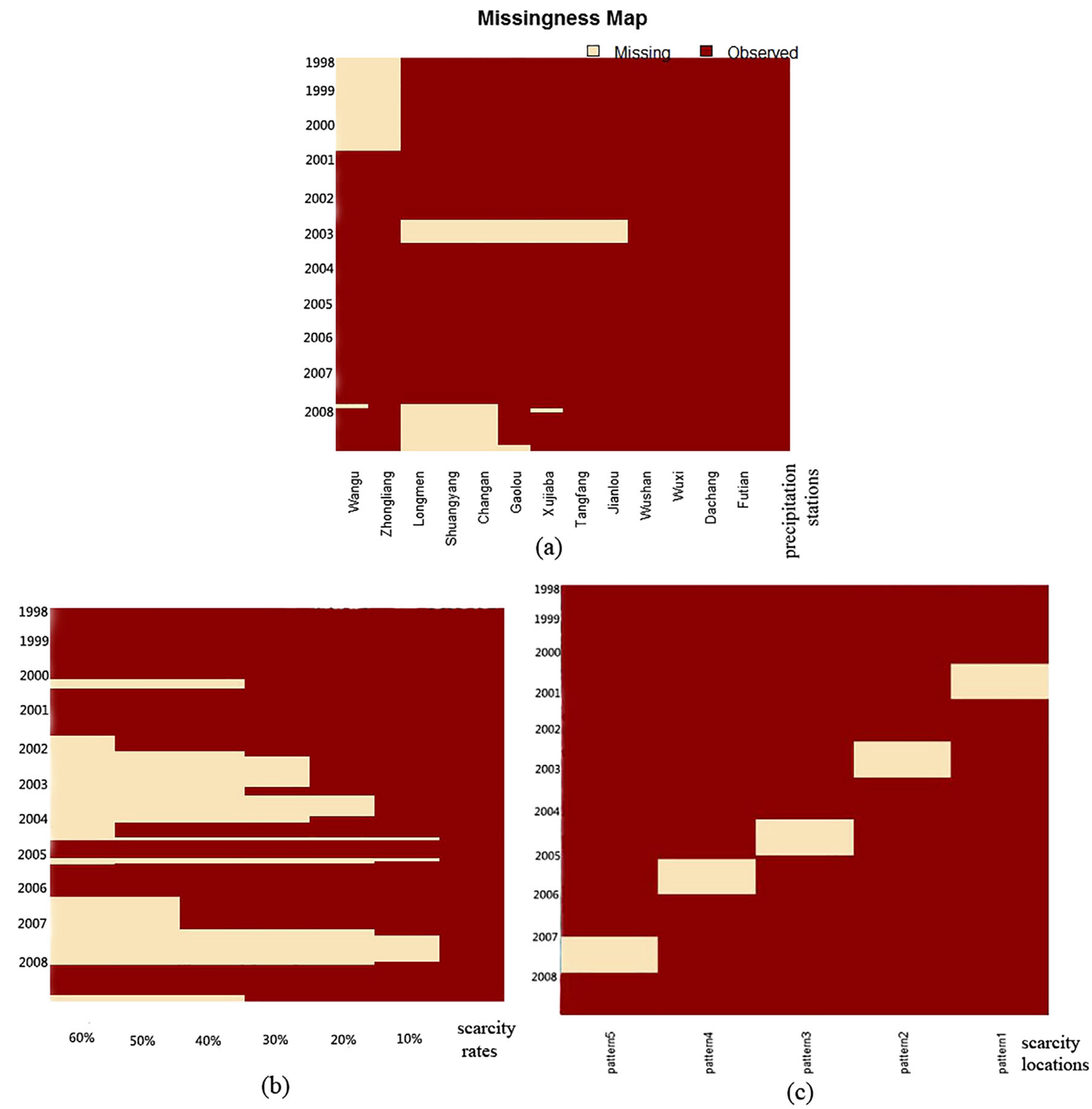

Fig. 2. Actual absence of rainfall data from the rainfall stations in the Daning River watershed and the daily rainfall data scarcity design for the Xining station Note: Fig. 2(a) represents the actual data scarcity in the Daning River watershed. And the daily rainfall data were from 1998 to 2008 were collected at nine rainfall stations inside the watershed and four gauges at sites approximately 30 km outside the watershed boundary. Fig. 2(b) represents the different scarcity rates design of daily rainfall data from 1998 to 2008 for the Xining station. Fig. 2(c) represents the different scarcity location design of daily rainfall data from 1998 to 2008 for the Xining station.

the simulated H/TP outputs changed dramatically from the perspective of the whole watershed.

The $\Delta H_{spatial}$ values for each sub-watershed are shown in Fig. 4. It can be seen the $\Delta H_{spatial}$ for the 7th sub-watershed and its other downstream sub-watersheds such as 9th sub-watershed increased during S1 scenario (scarcity of the Xining station). At the same time, as the total information entropy decreased, the sub-watersheds with greater $\Delta H_{spatial}$ increased, indicating information entropy could reflect changes in trends and processes compare to traditional evaluation methods employing Ens and R^2. In general, the absence of rainfall stations had different impacts on the flow and TP load simulations, while the Xining station is identified as the key rainfall station in this region. By comparing information entropy, the lack of rainfall stations containing critical amounts of information had a more impact on the

Table 2
Evaluation of the simulation effects for the different combinations of rainfall stations.

	Evaluation indicators	9-gauge	8-gauge	5-gauge	4-gauge	3-gauge
Flow	ENS	0.7425	0.5729	0.7308	0.6273	0.6023
	R^2	0.786	0.638	0.771	0.676	0.66
	ΔH		22.5170	9.1101	16.9243	25.4811
Total phosphorus	ENS	0.9299	0.6928	0.8299	0.7415	0.742
	R^2	0.952	0.867	0.947	0.943	0.94
	ΔH		32.4108	29.6263	37.0232	47.7475

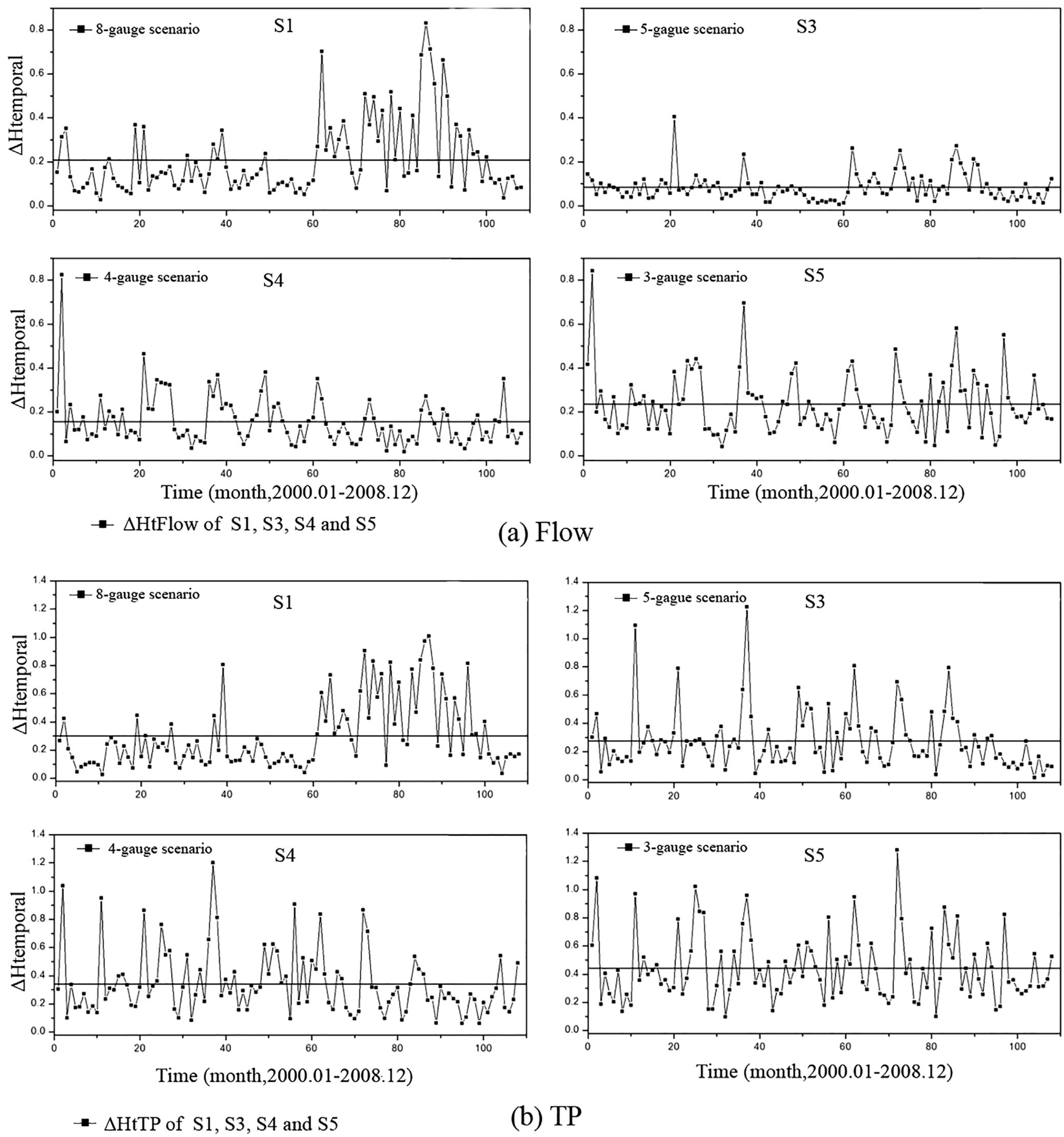

Fig. 3. $\Delta H_{temproal}$ for different combinations of rainfall stations. (a) Flow and (b) TP.

simulation results, indicating the lack of critical sites was more important than the reduction in the total amount of information contained in each scenario.

3.1.2. Impacts on the downstream simulations

In this section, the change of entropy information of downstream reaches during different scenarios was further taken into account. As mentioned above, the model performances for the S0 and S2 scenarios were identical due to the interpolation using Thiessen polygons. Seven stations were therefore chosen for this analysis and the $\Delta H_{spatial}$ values of each sub-watershed are shown in Fig. 5. It can be seen that the effects of the different missing rainfall stations on H/TP simulations were distinctly different. The Gaolou, Jianlou and Xining stations had large influences on the simulated flow and there were large $\Delta H_{spatial}$ in some sub-watersheds. If the data of Gaolou station were missing, the vast majority of the sub-watersheds had changes in entropy exceeding 0.6. When the data from the Jianlou station were missing, the 12th sub-watershed experienced a large change in entropy (exceeding 1.0). In addition to impacts of the Wuxi and Tangfang stations, the absence of other rainfall stations had greater impacts on the TP simulations. In particular, if the Xining station were omitted, almost all of the sub-watersheds experienced changes in entropy exceeding 0.2.

The attenuation of $\Delta H_{spatial}$ in downstream processes could be used to explain the impacts of data scarcity on the downstream simulations.

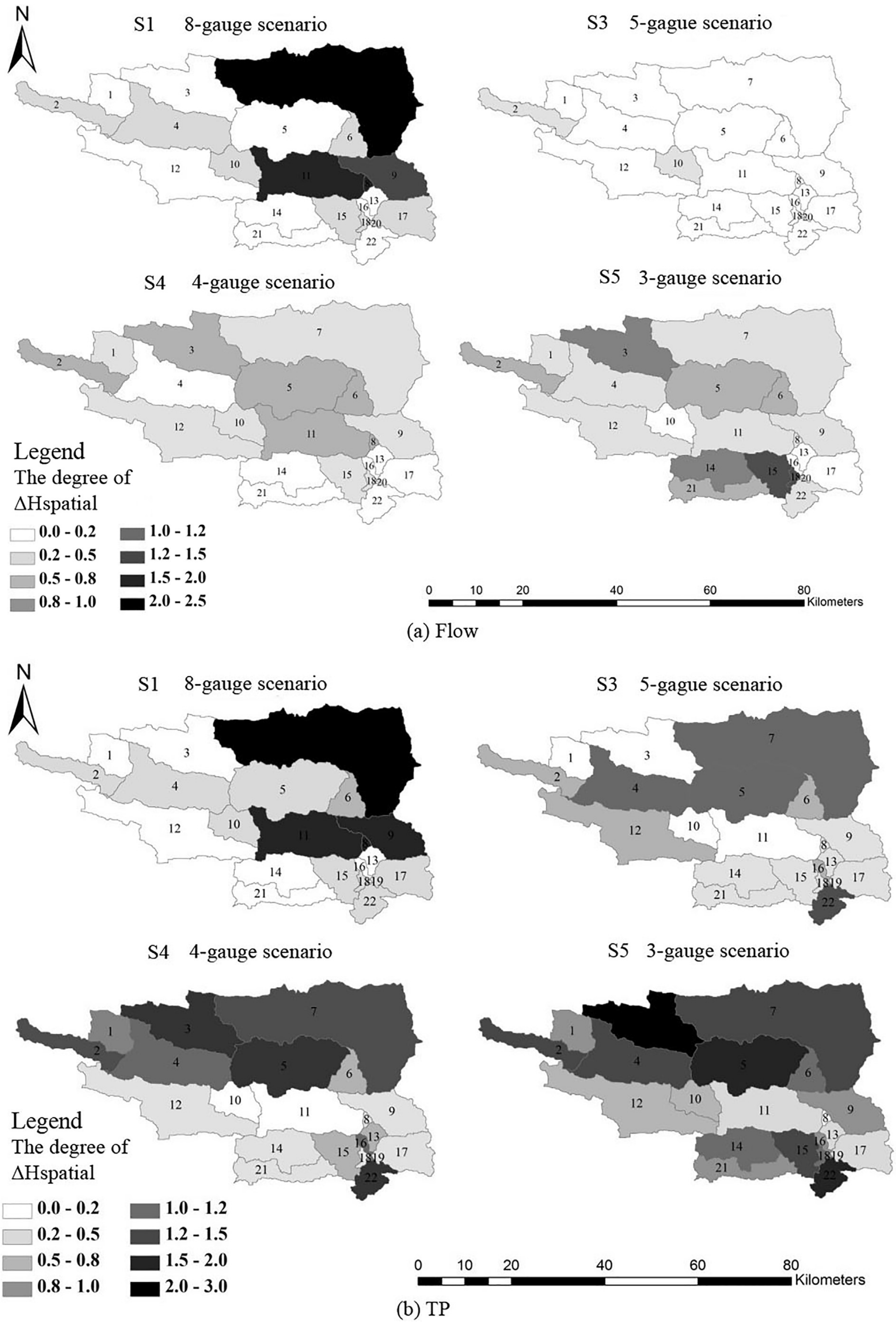

Fig. 4. $\Delta H_{spatial}$ for different combinations of rainfall stations. (a) Flow and (b) TP.

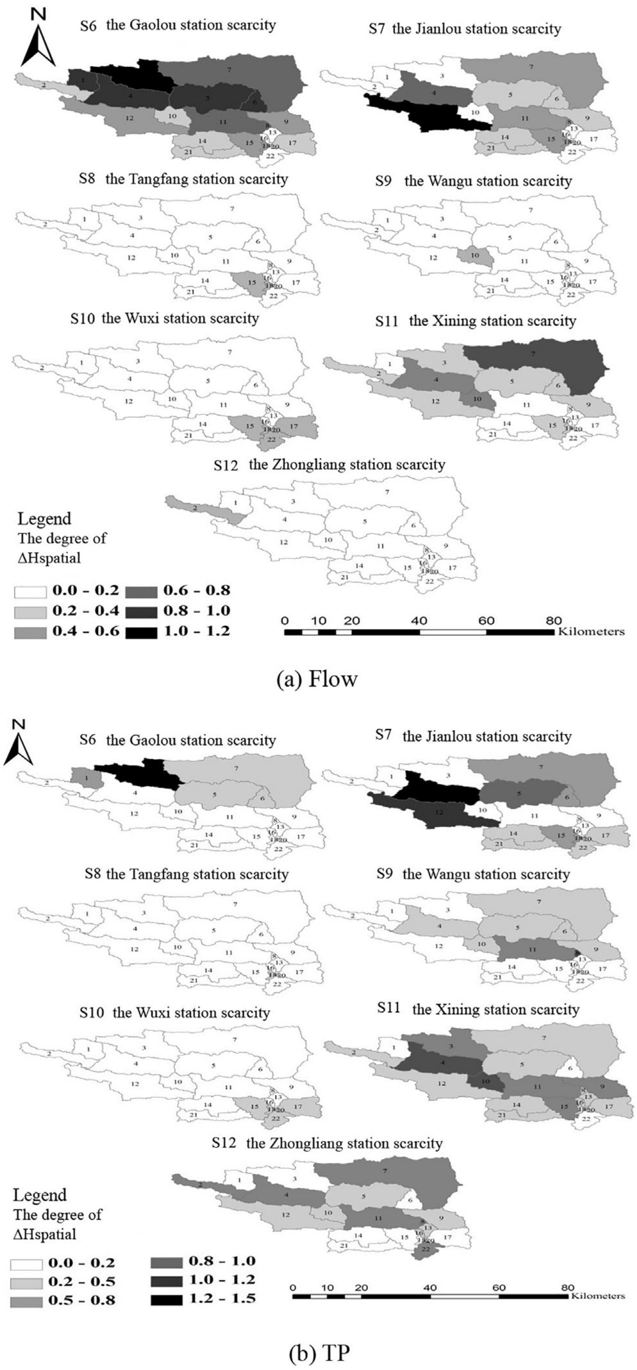

Fig. 5. $\Delta H_{spatial}$ for scarce data from each station. (a) Flow and (b) TP.

As shown in Fig. 1, these listed stations were on the associated tributaries (the Gaolou and Zhongliang stations on the Xixi River, the Jianlou and Wanguo stations on the Houxi River, the Tangfang and Wuxi stations on the Baiyang River, and the Xining station on the Dongxi River). For this paper, the four major tributaries were divided into eight small tributaries according to the different starting points. The attenuation process of ΔH for flow and TP load in the tributary process are shown in Fig. 6. Fig. 6(a) shows that although the trend was consistent, the attenuation process of ΔH in the Xixi River for the Gaolou station greatly exceeded that for the Zhongliang station. That is, in this tributary, the Gaolou station had a greater impact on the downstream simulations, indicating the scarcity of the Gaolou station was more likely to cause a larger deviation from the simulation results. Similarly, when compared with the Wangu station, the omission of the Jianlou station had a greater effect on the downstream simulated flows.

The attenuation of the ΔH for the Tangfang and Wuxi stations were consistent but as the Wuxi station is closer to the export of the watershed, the impact of Wuxi station was still large on the end of the tributary. Compared with that for flow, the attenuation process of ΔH for TP was relatively large, but the overall trend remained consistent. The Gaolou, Jianlou, Wuxi and Xining stations produced relatively large impacts on the simulations of each tributary due to the effects of information transmission across basins and because there are no corresponding rainfall stations in the sub-watershed near the Gaolou, Jianlou and Xining stations. In addition, SWAT is based on the rainfall interpolation method using Thiessen polygons so the rainfall data from those sites might be interpolated to other tributaries, showing the cross-tributary impacts.

In our previous study (Shen et al., 2012), it was clearly noted that the application of Thiessen polygons would cause uncertainties in rainfall spatial heterogeneities. Coincidentally, Dile and Srinivasan, 2015 and Yu et al., 2011 also gave the same conclusions. This was confirmed in more detail in this paper by considering the downstream and cross-tributary impacts. Thus, critical rainfall locations could be found by quantifying their impacts on the up-downstream simulations. Scarce data from key rainfall stations will cause simulation results to widely deviate from the actual situation.

3.2. Impacts of temporal data scarcity

3.2.1. Impacts of missing rates

In this section, the Xining station is used to quantify the impacts of different missing rates and different location scarcities due to its information entropy and impacts on downstream simulations. As shown in Table 3, the Ens results showed a downward trend as the missing rate increased. For missing rates exceeding 20%, the Ens were less than 0.6 and decreased continuously. Under the different missing rates, the TP simulation results were consistent with the flow results. For missing rates less than 60%, the Ens exceeded 0.8, indicating satisfactory TP simulations under these missing rates. However, Ens dropped to 0.75 if a missing rate reached 60% primarily due to the small amount of TP monitoring data. In this sense, if 60% of the rainfall data are missing, this would result in a poor point-to-point comparison between TP simulation and measured TP as this scarcity condition is a bit large and might cover most of the paired TP monitoring data. As a whole, these results showed that 60% is a critical threshold. Once this missing rate is exceeded, the simulation results will immediately become poor.

The $\Delta H_{temporal}$ for the different missing rates are shown in Fig. 7. It can be seen that there were large differences in $\Delta H_{temporal}$ for the flows and TP simulations in different months. In addition, $\Delta H_{temporal}$ increased with increasing missing rates. This indicates that a high missing rate of rainfall-driven data is likely to have a large effect on the H/TP simulations. At the same missing rate, the $\Delta H_{temporal}$ for flow and TP simulations were consistent with the change trend over the months, but the $\Delta H_{temporal}$ in TP was larger than for flow. That is, when compared to the flow, the simulation results of TP were more affected by the absence of rainfall data.

When the missing rate was gradually increased, the information entropy and Ens exhibited the same trends. It can be seen from Table 3 that when the rainfall-driven data loss rate exceeded 50%, the ΔH tended to be flat without further change. This critical threshold is consistent with the value mentioned above because as the missing rate increases, the mean rainfall data gradually approach the threshold. The weather generator also reaches its interpolation limit. At that time, the simulation results will have larger deviations.

3.2.2. Impacts of scarcity location

As shown in Fig. 8 and Table 2, the simulation results of flow and TP were greatly influenced by data missing positions and different location scarcities caused large differences. Fig. 8(a) shows that the simulated

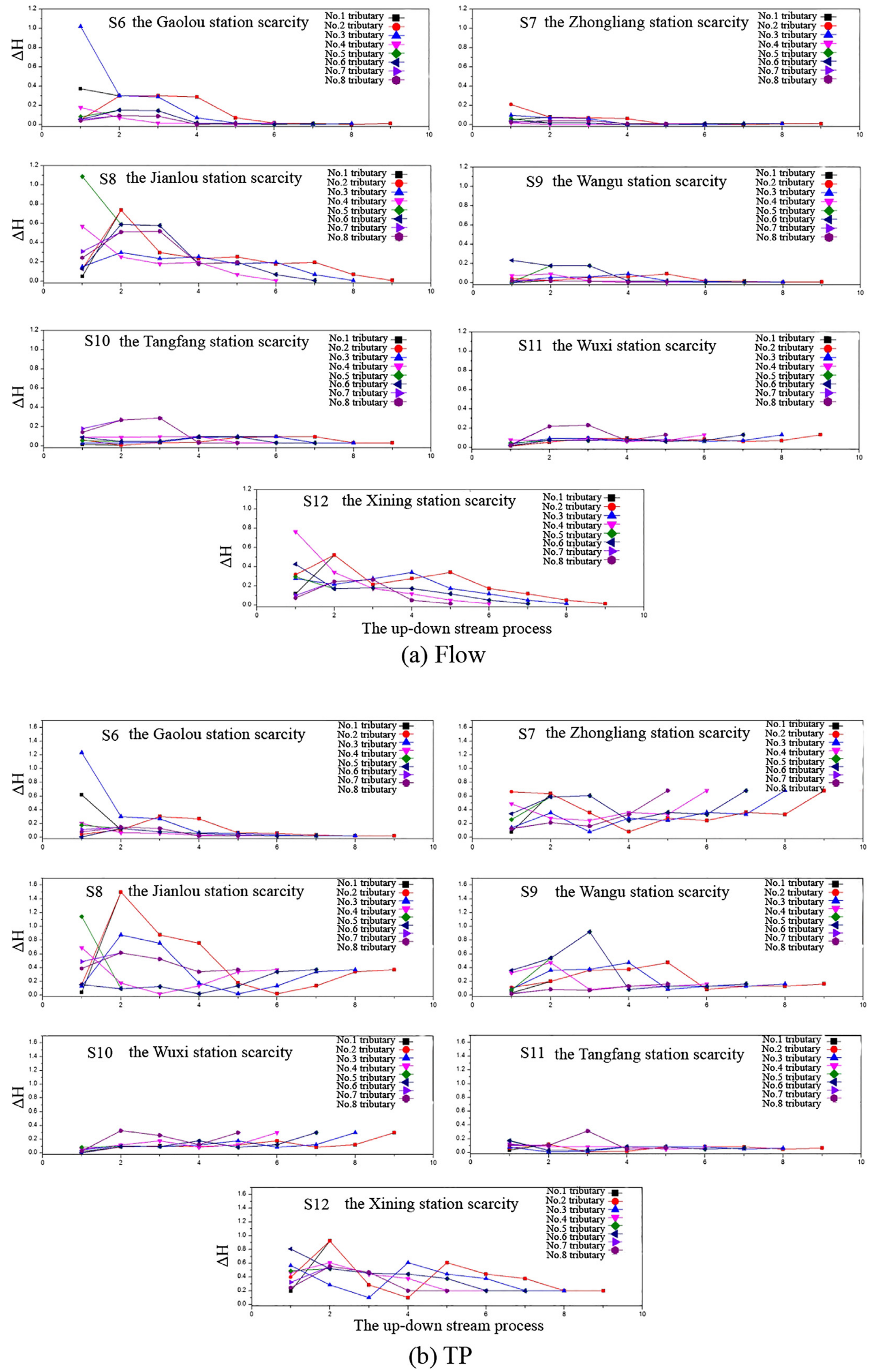

Fig. 6. Attenuation processes for ΔH in the upstream downstream process. (a) Flow and (b) TP. Note: The abscissa of Fig. 6 means that each tributary flows from upstream to downstream process. The No.1 tributary flows from 1th, 4th, 5th, 6th, 9th, 13th, 19th, 20th to 22th sub-watershed. The No. 2 tributary flows from 2th, 4th, 5th, 6th, 9th, 13th, 19th, 20th to 22th sub-watershed. The No. 3 tributary flows from 3th, 5th, 6th, 9th, 13th, 19th, 20th to 22th sub-watershed. The No. 4 tributary flows from 7th, 9th, 13th, 19th, 20th to 22th sub-watershed. The No. 5 tributary flows from 12th, 11th, 8th, 13th, 19th, 20th to 22th sub-watershed. The No. 6 tributary flows from 10th, 11th, 8th, 13th, 19th, 20th to 22th sub-watershed. The No. 7 tributary flows from 14th, 15th, 18th, 20th to 22th sub-watershed. The No. 8 tributary flows from 21th, 15th, 18th, 20th to 22th sub-watershed.

Table 3
Evaluation of the simulation results for the different temporal data scarcities.

		No missing	Different missing rates						Different data location scarcity				
			10%	20%	30%	40%	50%	60%	Pattern1	Pattern2	Pattern3	Pattern4	Pattern5
Flow	Ens	0.7425	0.6054	0.5457	0.5677	0.5366	0.5693	0.5494	0.6871	0.7287	0.7342	0.6573	0.6103
	R^2	0.786	0.662	0.611	0.649	0.612	0.641	0.633	0.741	0.778	0.787	0.714	0.667
	ΔH		3.7115	6.0821	8.2601	11.7208	14.3642	15.508	2.0351	2.5275	1.7184	3.6907	3.9754
TP	Ens	0.9299	0.8655	0.8711	0.8242	0.8615	0.8613	0.7515	0.9329	0.9137	0.9289	0.4339	0.8514
	R^2	0.952	0.933	0.932	0.905	0.928	0.914	0.794	0.954	0.934	0.961	0.551	0.925
	ΔH		5.9542	9.2635	11.5274	16.8487	20.6244	21.3764	4.0201	4.0038	3.0790	5.7503	6.5524

values of the missing segments in patterns 1, 4 and 5 had large deviations from the measured values. It can also be seen from Table 3 that the effects of the simulation on the flows for the three modes were worse than those for the other modes, primarily because patterns 1, 4 and 5 corresponded to 2000, 2005 and 2007, which were high flow years. This indicated missing data from high flow years can easily lead to large deviations in overall rainfall averages, which can import greater errors on data generated by the weather generator and measured rainfall data, thereby affecting overall simulation results. By using Ens values, Fig. 8(b) shows that the TP model performance was relatively good during different scarcity conditions. The simulated effect of the TP load for pattern 4 was significantly inferior to those for the other missing modes, primarily due to the absence of high values of TP data, which resulted in reduced Ens.

It can be seen from Fig. 8 that the flow and TP load simulations were not only affected by the positions with scarce data but also experienced certain impacts during the subsequent time period. Therefore, to investigate the effects of the SWAT model on the simulation results in the follow-up periods, the simulation results for the missing segments were analysed separately from that for the complete data state. The results show that there were significant differences in the model performances of flow for the different missing modes. The simulation results for normal flow years and low flow years (patterns 2 and 3) were generally better than those for the high flow years (patterns 4 and 5). The simulation results of TP for the different missing modes were consistent with those for flow rate; that is, the model performances for normal flow years were better than for high flow years. When compared to flow, however, the variation of the Ens for TP was greater. In addition, different missing patterns had certain impacts on the simulations during subsequent periods. For patterns 4 and 5, which represented data missing scenarios in high flow years, whether for flow or TP, the model performances of the missing segments had greater impacts on the follow-up simulations. The changes were larger, and the levels during normal flow and low flow years were relatively flat.

The monthly $\Delta H_{temporal}$ for flow and TP under different missing positions are shown in Fig. 9. The $\Delta H_{temporal}$ increased if data in high flow years were missing, and the model performance become poor, which indicates that missing positions had greater impacts on the simulation results, especially for missing positions in high flow years. When simulating TP, as shown in Fig. 9(b), there was a significant fluctuation in the $\Delta H_{temporal}$ for TP simulation after the year with missing data. At the same time, the $\Delta H_{temporal}$ for TP was obviously greater than the $\Delta H_{temporal}$ for flow rate, which shows that the absence of different data affected TP more than flow.

4. Discussion

Rainfall is one of the causes of NPS pollution and is inevitable as an important error source of H/NPS simulation (Ning et al., 2006; Rathnayake and Tanyimboh, 2015; Schreiber and Mcdowell, 1985). Based on the results, it could be concluded that rainfall data scarcity had a great impact on H/NPS simulation so it is necessary to consider the effects of data scarcity at the early stage of model setup, especially for those DHMs/SDHMs. This paper provided some practical implications for H/NPS simulation, as well as the optimal design of rainfall gauge. Due to limited financial and material resources, it is impossible to set up rainfall stations at each sub-watershed to capture actual rainfall spatial variability (Karimi-Hosseini et al., 2011; Lakshmi, 2016; Sombroek, 2001). This paper indicated that rainfall data at key sites in the large watershed are very important and will affect model performances greatly when such stations were scarce. Therefore, the importance of critical rainfall station is highlighted, even exceeding the importance of the number of rainfall stations. The critical-site rainfall stations are those stations that have the greatest impacts on the spatial heterogeneity of rainfall (Ali et al., 2005; Cheng et al., 2017; Swain and Patra, 2017). And it can be identified by comparing the changes of entropy values of simulated H/NPS during different scarcity scenarios. Besides, it should be noted that critical-site stations might be different for flow and NPS prediction because pollutants are influenced by more factors. For the Daning River watershed, the Xining station was identified as a key site as it is located the center of the catchment instead of the edge part. Such sites should be more easily to be focused on as their data might be interpolated into several sub-watersheds that were adjacent to them. Moreover, from the perspective from the downstream impacts, the rainfall station that near the catchment outlet is more likely to become a key station and has a major impact on the end of this tributary. Besides, the application of radar quantitative precipitation estimation to hydrology and water quality models can be preferred to interpolated rainfall point measurements because of the wide coverage that radars can provide, together with a good spatio-temporal resolutions so radar-rainfall measurements have been improving over the years. In those remote regions, due to unsatisfactory spatial interpolation of scarce rainfall data (Bhowmik and Costa, 2015), the use of the multi-sources of radar data and station data is further recommended for a more accurate description of rainfall spatial heterogeneity.

On the other hands, it can be concluded from above temporal series analysis that: (i) certain threshold values exist for missing rates; model performance might be stable within this threshold value by using the weather generator but once that threshold value is reached, model performance become worsen immediately. This indicated that small amount of scarce data can be interpolated by the weather generator, but the accuracy of the simulation changes rapidly when a scarcity threshold is exceeded. In the study area, 60% is a critical value for missing rates and data scarcity not exceed 60% of missing rate could ensure satisfactory model performances. In reality, the scarce data are inevitable due to a variety of human and natural factors (Wambura et al., 2017), but determining the threshold of input data can be helpful in clarifying the robustness of simulations and in reducing unnecessary monitoring costs. (ii) Compared to normal flow or low flow years, scarce rainfall data in high flow years will result in greater errors in the simulation results. Smith et al. (2016) proposed that scarce data during spring freshets will introduce greater uncertainty to the flow simulation. Thus, simulation results during those periods can be improved by explicitly indicating soil freezing. As for this paper, we recommend that rainfall data integrity in high flow periods should be ensured, especially for NPS prediction. And in actual monitoring, the monitor should try to

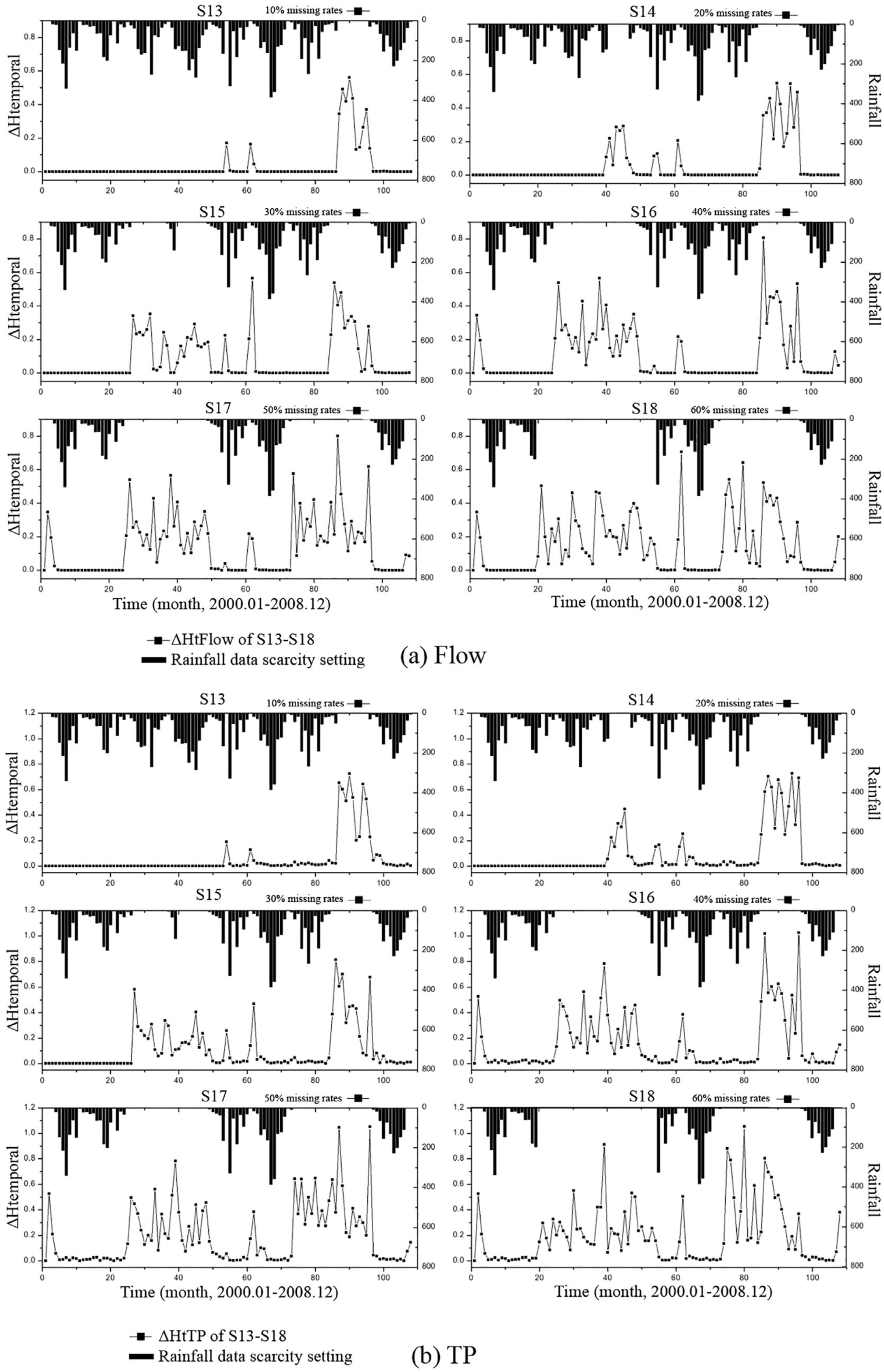

Fig. 7. The impact of different missing rates on H/NPS prediction. (a) Flow and (b) TP.

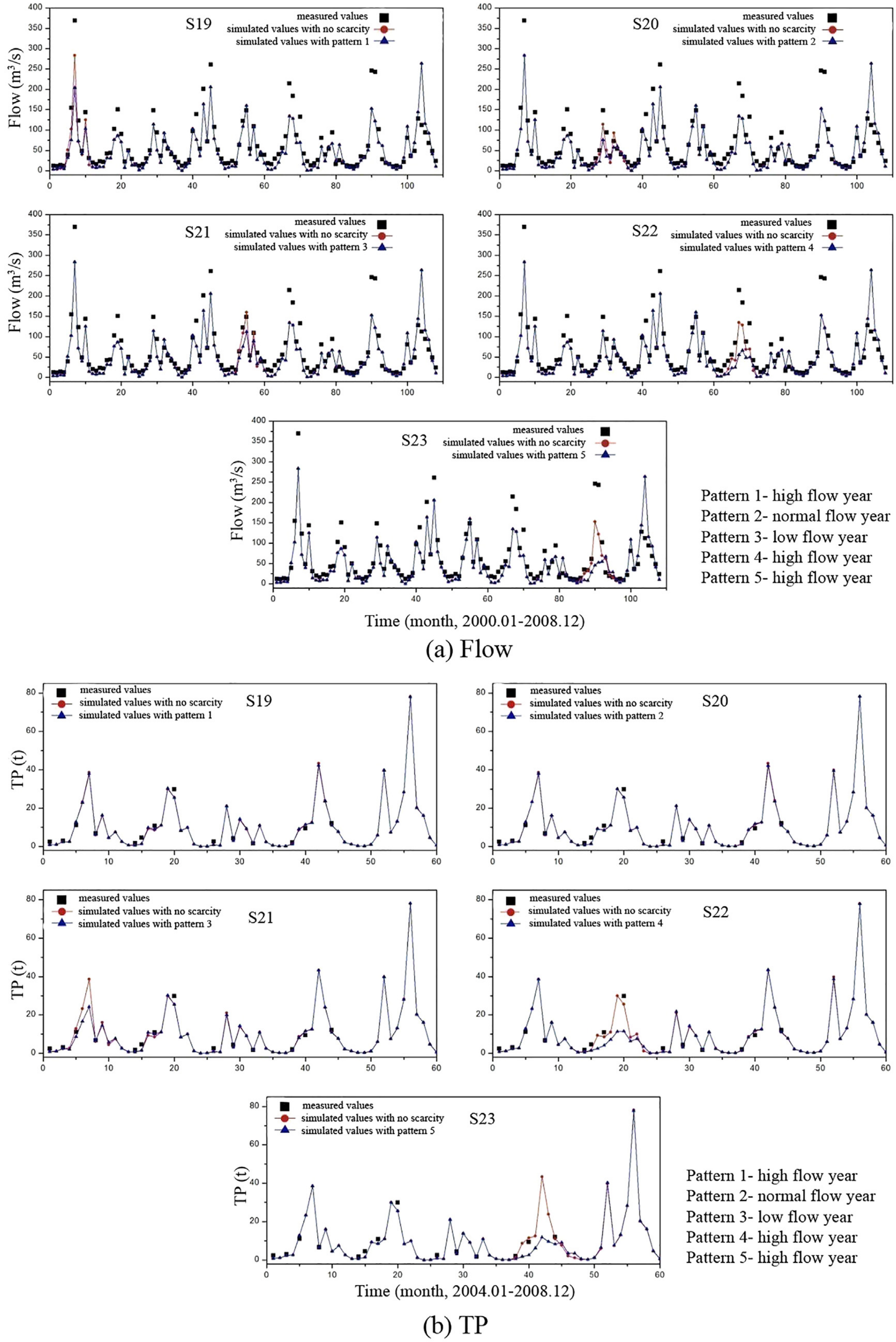

Fig. 8. Simulation results at different scarcity locations. (a) Flow and (b) TP.

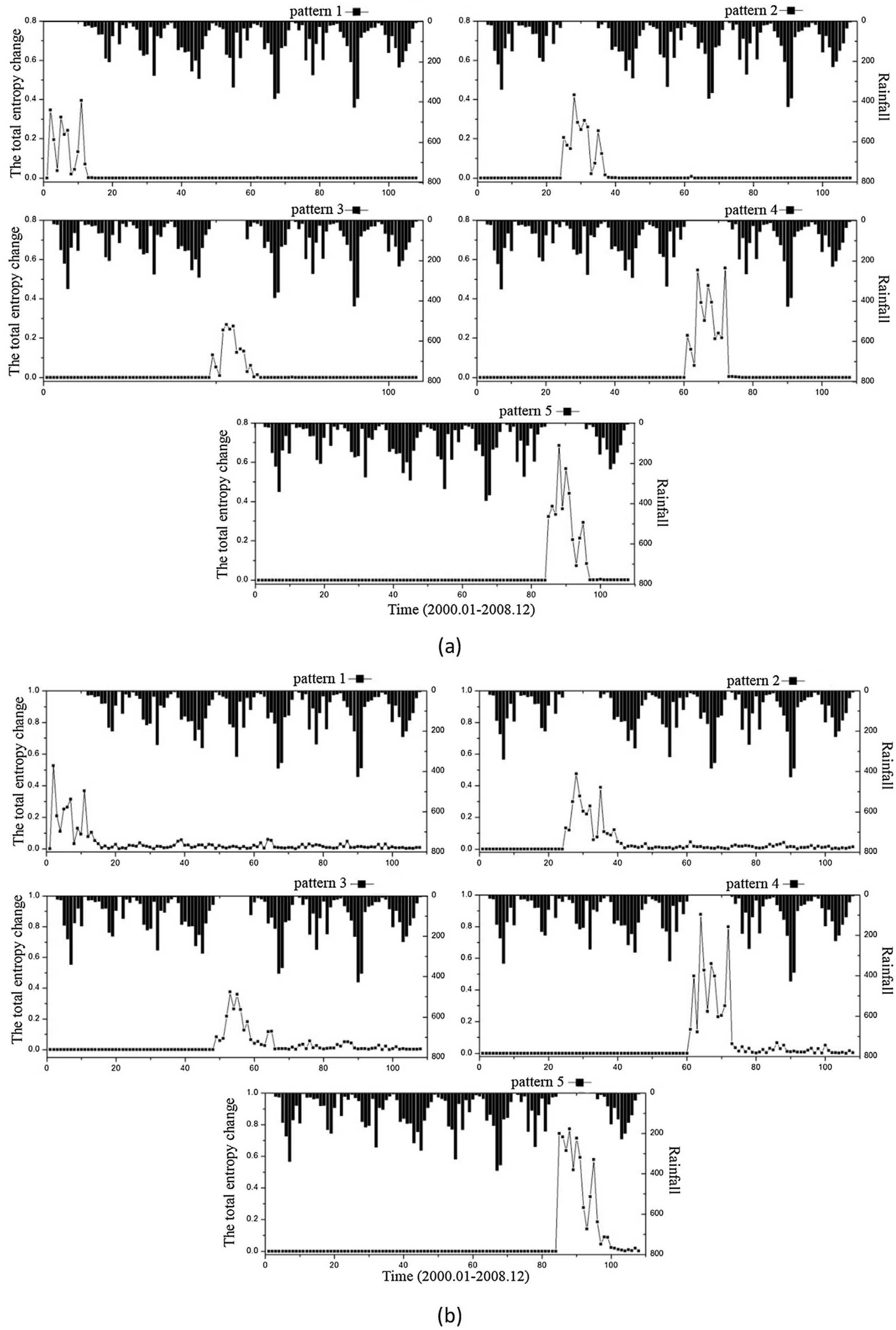

Fig. 9. $\Delta H_{temproal}$ of simulation results for different scarcity locations. (a) Flow and (b) TP.

ensure the integrity of rainfall data of the critical-site rainfall stations in the high flow years.

Another major task of this research is to introduce the information entropy method to evaluate the impacts of data scarcity on the H/NPS predictions. As mentioned above, the impacts of rainfall scarcity on the simulated H/NPS data should be considered not only at the catchment outlet (traditional indicators) but also theirs temporal-spatial distributions among sub-watersheds through entropy values. As shown in Fig. 8 and Table 2, the Ens values and information entropy were reduced from 0.7425 to 0.6023 and from 1.9145 to 0.8804, respectively. This indicated that although data scarcity might show little impact on traditional indicator, ΔH would show a more sensitive response to the rainfall input and higher ΔH reflected more impacts of data scarcity on simulation results. This is because the information entropy could reflect the impacts of data scarcity on the simulated data of each sub-watershed so this indicator outperforms traditional indicators during data scarcity scenarios. It can be seen from Fig. 9 that the total entropy for the TP load in different missing modes was nearly twice the total entropy of the flows, indicating the rainfall data scarcity would show more impacts on NPS-TP simulation. This may be due to a variety of factors affecting the P transport process, such as underlying surface, altitude and aspect (Sharpley and Syers, 1979). Therefore, the use of the combination of the information entropy method and Ens is suggested to evaluate the H/NPS models when the data has scarcity. Besides, it should be noted that the loss potentials of multiple pollutants are not evenly distributed on the same spatial and temporal scales so more studies should be conducted to provide more comprehensive conclusion.

5. Conclusions

In this study, the impacts of rainfall data scarcity on H/NPS simulations were quantified by running and comparing different spatial-temporal data scarcity scenarios. The results highlighted the importance of critical-site rainfall stations (Xining station in this paper) on the H/NPS simulations, which showed influences on ΔH, and have greater downstream and cross-tributary impacts. Certain threshold value of data scarcity rate (60% in this paper) did exist, beyond which the model preformation would become very poor. Compared to traditional indicators, information entropy could serve as a good substitute for the Ens because it reflect information at the sub-watershed scale and attention of impacts within river network. This paper provides important implications for the application of DHMs/SDHMs tools, as well as for the optimal design of rainfall gauges, especially in large basins.

However, the results of this study should be extended to other catchments with caution. For example, besides rainfall inputs, other parameters like fertilizer input would influence the H/NPS simulation, while weather generator methodology, which is used to construct some missing data in this study, also varies from model to model. Further studies are needed by considering the integrated impact of temporal and spatial data scarcity. In addition, more studies should also take the synergism of satellite remote sensing, Climate Forecast System Reanalysis (CFSR) product and station data especially in an extreme hydrological event.

Acknowledgements

This research was funded by the National Natural Science Foundation of China (Nos. 51579011 and 51779010), the Newton Fund (Grant Ref: BB/N013484/1), Key Laboratory of Nonpoint Source Pollution Control, Ministry of Agriculture, P.R. China (1610132016005) and the Interdiscipline Research Funds of Beijing Normal University.

References

Abbaspour, K.C., Vejdani, M., Haghighat, S., 2007. SWAT-CUP calibration and uncertainty programs for SWAT. Modsim International Congress on Modelling & Simulation Land Water & Environmental Management Integrated Systems for Sustainability. 364 (3), 1603–1609.
Ali, A., Lebel, T., Amani, A., 2005. Rainfall estimation in the Sahel. Part I: error function. J. Appl. Meteorol. 44 (11), 1691–1706.
Amatya, D.M., Jha, M., Edwards, A.E., Williams, T.M., Hitchcock, D.R., 2011. SWAT-based streamflow and embayment modeling of karst-affected chapel branch watershed, South Carolina. Trans. ASABE 54 (4), 1311–1323.
Andréassian, V., Perrin, C., Michel, C., Usart-Sanchez, I., Lavabre, J., 2001. Impact of imperfect rainfall knowledge on the efficiency and the parameters of watershed models. J. Hydrol. 250 (1), 206–223.
Bardossy, A., Plate, E.J., 1992. Space-time model for daily rainfall using atmospheric circulation patterns. Water Resour. Res. 28 (5), 1247–1259.
Bera, M., Borah, D.K., 2003. Watershed-scale hydrologic and non-point source pollution models: review of mathematical bases. Trans. ASABE 46 (6), 1553–1566.
Bhowmik, A.K., Costa, A.C., 2015. Representativeness impacts on accuracy and precision of climate spatial interpolation in data-scarce regions. Meteorol. Appl. 22 (3), 368–377.
Bian, Y.W., Feng, Y., 2010. Resource and environment efficiency analysis of provinces in China: a DEA approach based on Shannon's entropy. Energy Policy 38 (4), 1909–1917.
Bieger, K., Hörmann, G., Fohrer, N., 2014. Simulation of streamflow and sediment with the soil and water assessment tool in a data scarce catchment in the Three Gorges Region, China. J. Environ. Quality 43 (1), 37.
Biggs, E.M., Atkinson, P.M., 2015. A comparison of gauge and radar precipitation data for simulating an extreme hydrological event in the Severn Uplands, UK. Hydrol. Process. 25 (5), 795–810.
Chaubey, I., Haan, C.T., Salisbury, J.M., et al., 1999. Quantifying model output uncertainty due to the spatial variability of rainfall. J. Am. Water Resour. Assoc. 35 (5), 1113–1123.
Chen, L., Shen, Z., Yang, X., Liao, Q., Yu, S.L., 2014. An Interval-Deviation Approach for hydrology and water quality model evaluation within an uncertainty framework. J. Hydrol. 509 (4), 207–214.
Chen, L., Sun, C., Wang, G., Xie, H., Shen, Z., 2017a. Event-based nonpoint source pollution prediction in a scarce data catchment. J. Hydrol. 552, 13–27.
Chen, L., Sun, C., Wang, G., Xie, H., Shen, Z., 2017b. Modeling multi-event non-point source pollution in a data-scarce catchment using ANN and entropy analysis. Entropy 19 (6), 265.
Cheng, M., et al., 2017. Performance assessment of spatial interpolation of precipitation for hydrological process simulation in the three gorges basin. Water 9 (11), 838.
Cho, J., Bosch, D., Lowrance, R., Strickland, T., Vellidis, G., 2009. Effect of spatial distribution of rainfall on temporal and spatial uncertainty of SWAT output. Trans. ASABE 52 (5), 277–281.
Cluckie, I.D. et al., 2004. Weather radar technology and future developments., International Conference of GIS and Remote Sensing in Hydrology, Water Resources and Environment (ICGRSHWE), Three Gorges Dam, China. 16-19 September 2003.
Dile, Y.T., Srinivasan, R., 2015. Evaluation of CFSR climate data for hydrologic prediction in data-scarce watersheds: an application in the Blue Nile River Basin. J. Am. Water Resour. Assoc. 50 (5), 1226–1241.
Douglas-Mankin, K.R., Srinivasan, R., Arnold, A.J., 2010. Soil and Water Assessment Tool (SWAT) Model: Current Developments and Applications. Trans. ASABE 53 (5), 1423–1431.
Drecht, G.V., Bouwman, A.F., Knoop, J.M., et al., 2003. Global modeling of the fate of nitrogen from point and nonpoint sources in soils, groundwater, and surface wate. Global Biogeochem. Cycles 17 (4) 10.1029/2003GB00 2060.
Fontaine, T.A., Cruickshank, T.S., Arnold, J.G., Hotchkiss, R.H., 2002. Development of a snowfall–snowmelt routine for mountainous terrain for the soil water assessment tool (SWAT). J. Hydrol. 262 (1–4), 209–223.
Gong, Y., Shen, Z., Liu, R., Qian, H., Wu, X., 2012. A comparison of single- and multi-gauge based calibrations for hydrological modeling of the Upper Daning River Watershed in China's Three Gorges Reservoir Region. Hydrol. Res. 43 (6), 822–832.
Harrison, D.L., Kitchen, M., Scovell, R.W., 2009. High-resolution precipitation estimates for hydrological uses. Water Manage. 162 (2), 125–135.
Hartkamp, A.D., White, J.W., Hoogenboom, G., 2003. Comparison of three weather generators for crop modeling: a case study for subtropical environments. Agric. Syst. 76 (2), 539–560.
Heathwaite, A.L., Dils, R.M., 2000. Characterising phosphorus loss in surface and subsurface hydrological pathways. Sci. Total Environ. 251, 523–538.
Hrachowitz, M., et al., 2013. A decade of Predictions in Ungauged Basins (PUB)—a review. Hydrol. Sci. J. 58 (6), 1198–1255.
Hrachowitz, M., Fovet, O., Ruiz, L., Savenije, H.H.G., 2016. Transit time distributions, legacy contamination and variability in biogeochemical 1/fα scaling: how are hydrological response dynamics linked to water quality at the catchment scale? Hydrol. Process. 29 (25), 5241–5256.
Hrachowitz, M., Clark, M.P., 2017. HESS Opinions: the complementary merits of competing modelling philosophies in hydrology. Hydrol. Earth Syst. Sci. Discuss. 21 (8), 1–22.
Ilunga, M., 2017. Assessing catchment resilience using entropy associated with mean annual runoff for the upper vaal catchment in South Africa. Entropy 19 (5), 147.
Karimi-Hosseini, A., Haddad, O.B., Mariño, M.A., 2011. Site selection of raingauges using entropy methodologies. Water Manage. 164 (7), 321–333.
Kashani, M.H., Ghorbani, M.A., Dinpashoh, Y., Shahmorad, S., 2016. Integration of

Volterra model with artificial neural networks for rainfall-runoff simulation in forested catchment of northern Iran. J. Hydrol. 540, 340–354.
Khosravi, K., Pourghasemi, H.R., Chapi, K., Bahri, M., 2016. Flash flood susceptibility analysis and its mapping using different bivariate models in Iran: a comparison between Shannon's entropy, statistical index, and weighting factor models. Environ. Monit. Assess. 188 (12), 656.
Kilsby, C.G., et al., 2007. A daily weather generator for use in climate change studies. Environ. Modell. Software 22 (12), 1705–1719.
Kuangyao, L., Fisher, T.R., Jordan, T.E., Correll, D.L., Weller, D.E., 2000. Modeling the hydrochemistry of the Choptank River Basin using GWLF and arc/info: 1. Model calibration and validation. Biogeochemistry 49 (2), 143–173.
Lakshmi, V., 2004. The role of satellite remote sensing in the Prediction of Ungauged Basins. Hydrol. Process. 18 (5), 375–387.
Lakshmi, V., 2016. Beyond GRACE: using satellite data for groundwater investigations. Ground Water 54 (5), 615–618.
Legates, D.R., McCabe, G.J., 1999. Evaluating the use of “goodness-of-fit” Measures in hydrologic and hydroclimatic model validation. Water Resour. Res. 35 (1), 233–241.
Lobligeois, F., Andréassian, V., Perrin, C., Tabary, P., Loumagne, C., 2014. When does higher spatial resolution rainfall information improve streamflow simulation? An evaluation on 3620 flood events. Hydrol. Earth Syst. Sci. Discuss. 18 (10), 12485–12536.
Lopez, M.G., Seibert, J., 2016. Influence of hydro-meteorological data spatial aggregation on streamflow modelling. J. Hydrol. 541, 1212–1220.
Merz, R., Blöschl, G., 2004. Regionalisation of catchment model parameters. J. Hydrol. 287 (1), 95–123.
Meusburger, K., Steel, A., Panagos, P., Montanarella, L., Alewell, C., 2012. Spatial and temporal variability of rainfall erosivity factor for Switzerland. Hydrol. Earth Syst. Sci. 16 (1), 167–177.
Michaelides, S.C., Tymvios, F.S., Michaelidou, T., 2009. Spatial and temporal characteristics of the annual rainfall frequency distribution in Cyprus. Atmos. Res. 94 (4), 606–615.
Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I-A discussion of principles. J. Hydrol. 10 (3), 282–290.
Ning, S.K., Chang, N.B., Jeng, K.Y., Tseng, Y.H., 2006. Soil erosion and non-point source pollution impacts assessment with the aid of multi-temporal remote sensing images. J. Environ. Manage. 79 (1), 88–101.
Pereiracardenal, S.J., Riegels, N.D., Berry, P.A., Smith, R.G., Yakovlev, A., Siegfried, T.U., Bauer-Gottwein, P., 2011. Real-time remote sensing driven river basin modeling using radar altimetry. Hydrol. Earth Syst. Sci. 15 (1), 241–254.
Rathnayake, U.S., Tanyimboh, T.T., 2015. Evolutionary multi-objective optimal control of combined sewer overflows. Water Resour. Manage. 29 (8), 2715–2731.
Richardson, C.W., 1981. Stochastic simulation of daily precipitation, temperature, and solar radiation. Water Resour. Res. 17 (1), 182–190.
Ruan, H., Zou, S.B., Yang, D.W., Wang, Y.H., Yin, Z.L., Lu, Z.X., Li, F., Xu, B.R., 2016. Runoff simulation by SWAT model using high-resolution gridded precipitation in the upper Heihe River Basin, Northeastern Tibetan Plateau. Hydrol. Earth Syst. Sci. Discuss. 9 (11), 1–23.
Schreiber, J.D., Mcdowell, L.L., 1985. Leaching of nitrogen, phosphorus, and organic-carbon from wheat straw residues. 1. Rainfall Intensity. J. Environ. Qual. 14 (2), 251–256.
Sharpley, A.N., Mcdowell, R.W., Kleinman, P.J.A., 2001. Phosphorus loss from land to water: integrating agricultural and environmental management. Plant Soil 237 (2), 287–307.
Sharpley, A.N., Syers, J.K., 1979. Effect of aerial topdressing with superphosphate on the loss of phosphate from a pasture catchment. N. Z. J. Agric. Res. 22 (2), 273–277.
Shemshadi, A., Shirazi, H., Toreihi, M., Tarokh, M.J., 2011. A fuzzy VIKOR method for supplier selection based on entropy measure for objective weighting. Expert Syst. Appl. 38 (10), 12160–12167.
Shen, Z., Chen, L., Liao, Q., Liu, R., Hong, Q., 2012. Impact of spatial rainfall variability on hydrology and nonpoint source pollution modeling. J. Hydrol. 472–473 (24), 205–215.
Shen, Z., Xie, H., Chen, L., Qiu, J., Zhong, Y., 2015a. Uncertainty analysis for nonpoint source pollution modeling: implications for watershed models. Int. J. Environ. Sci. Technol. 12 (2), 739–746.
Shen, Z., Zhong, Y., Huang, Q., Chen, L., 2015b. Identifying non-point source priority management areas in watersheds with multiple functional zones. Water Res. 68, 563–571.
Singh, P.K., Cherukuri, A.K., Li, J., 2014. Concepts reduction in formal concept analysis with fuzzy setting using Shannon entropy. Int. J. Mach. Learn. Cybern. 6 (1), 1–11.
Smil, V., 2000. Phosphorus in the environment: natural flows and human interferences. Annu. Rev. Energy Env. 25 (1), 53–88.
Smith, A., Welch, C., Stadnyk, T., 2016. Assessment of a lumped coupled flow-isotope model in data scarce Boreal catchments. Hydrol. Process. 30 (21), 3871–3884.
Sombroek, W., 2001. Spatial and temporal patterns of Amazon rainfall. Consequences for the planning of agricultural occupation and the protection of primary forests. Ambio 30 (7), 388–396.
Sun, W., et al., 2017. Physically based distributed hydrological model calibration based on a short period of streamflow data: case studies in four Chinese basins. Hydrol. Earth Syst. Sci. 21, 1–20.
Swain, J.B., Patra, K.C., 2017. Streamflow estimation in ungauged catchments using regionalization techniques. J. Hydrol. 554, 420–433.
Troin, M., Vallet-Coulomb, C., Piovano, E., Sylvestre, F., 2012. Rainfall–runoff modeling of recent hydroclimatic change in a subtropical lake catchment: Laguna Mar Chiquita, Argentina. J. Hydrol. 475 (6), 379–391.
Viviroli, D., Zappa, M., Gurtz, J., Weingartner, R., 2009. An introduction to the hydrological modelling system PREVAH and its pre- and post-processing-tools. Environ. Modell. Software 24 (10), 1209–1222.
Wambura, F.J., Dietrich, O., Lischeid, G., 2017. Evaluation of spatio-temporal patterns of remotely sensed evapotranspiration to infer information about hydrological behaviour in a data-scarce region. Water 9 (5), 333.
Wang, G., Chen, L., Huang, Q., Xiao, Y., Shen, Z., 2016. The influence of watershed subdivision level on model assessment and identification of non-point source priority management areas. Ecol. Eng. 87, 110–119.
Wilks, D.S., 2009. A gridded multisite weather generator and synchronization to observed weather data. Water Resour. Res. 45 (10), 2381–2386.
Yu, M., Chen, X., Li, L., et al., 2011. Streamflow simulation by SWAT using different precipitation sources in large arid basins with scarce raingauges. Water Resour. Manage. 25 (11), 2669–2681.
Zhang, X., Srinivasan, R., Debele, B., Hao, F., 2008. Runoff simulation of the headwaters of the yellow river using The SWAT model with three snowmelt algorithms 1. J. Am. Water Resour. Assoc. 44 (1), 48–61.