Temporal and spatial transferabilities of hydrological models under different climates and underlying surface conditions

Wushuang Yang[1, 2], Hua Chen[1, 2, 4*], Chong-Yu Xu[3*], Ran Huo[1, 2], Jie Chen[1, 2], Shenglian Guo[1, 2]

[1]State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan 430072, China

[2]Hubei Provincial Key Lab of Water System Science for Sponge City Construction, Wuhan University, Wuhan, China

[3]Department of Geosciences, University of Oslo, P O Box 1047 Blindern, N-0316 Oslo, Norway

[4]Department of Geography and Geographic Information Science, University of Illinois at Urbana-Champaign, 61801, the United States.

*Chua@whu.edu.cn; chongyu.xu@geo.uio.no

Abstract

Changing conditions of the climate and underlying surface have altered the rainfall-runoff relationships in many basins, greatly increasing additional challenges in the applicability of hydrological models for studying the hydrological response to those potential changes. However, systematic and simultaneous testing and comparing of both temporal and spatial transferabilities of different hydrological models under changing conditions have not received enough attention. The present study investigates the potential differences between temporal and spatial transferabilities of different hydrological models under different climatic and underlying surface conditions, which are synthesized from two basins in Southern China with 50-year historical records (1966–2015). The transferability of five hydrological models, i.e., XAJ, HBV, SIMHYD, IHACRES and GR4J, is investigated under the synthesised changing conditions by using a new evaluation method, proposed in this study. The results show that: (1) the proposed evaluation method is proved to be effective in evaluating the transferability of the models; (2) for temporal transferability under stationary condition, the five models show similar performances, but for spatial transferability, the performances of complex models (XAJ and HBV) are better than that of the simple model (GR4J); (3) the difference in underlying surface conditions in the target basin affects spatial transferability of the models; (4) hydrological models have much better transferability from dry to wet period than otherwise. This study provides an insight to test temporal and spatial transferabilities of hydrological models in the context of changing climate and underlying surface conditions.

Keywords: hydrological model, changing condition, spatial-temporal transferability, validation method

1. Introduction

The global climate and land use changes caused by substantial anthropogenic activities affect regional rainfall-runoff relationships, directly affecting local water resource availability (Arnell, 2004; Frich et al., 2002; Lu and Qin, 2020; Ma et al., 2008; Ragettli et al., 2020; Ye et al., 2013; Zhang et al., 2011, 2012). Scientific and accurate assessments of future water resources under changing environment have attracted more attention than before because water-related issues, such as flooding, drought and pollution, are becoming increasingly grave due to the impact of global warming and human activities (Alcamo et al., 2007; Chen et al., 2019; Doll, 2002; Li et al., 2015; Milly et al., 2008; Xiong et al., 2019). Hydrological models are the most important tool to study the impact of the changing environment on water resources (Chen et al., 2019; Fan et al., 2019; Guo et al., 2019; Xu and Singh, 2004). Hydrological models have several advantages in studying the impact of environment change (Gleick, 1986; Jiang et al., 2007; Klemes, 1986; Schulze, 1997). Firstly, many models are already available for different climatic or physiographic conditions, increasing flexibility in identifying and choosing the most appropriate model to evaluate any specific region. Secondly, extensive climate change scenarios obtained by climate models can be used as inputs for hydrological models when assessing the hydrological response to climate change. Thirdly, hydrological models are easy to manipulate and improve for specific areas or conditions. They are usually calibrated by using historic records, assuming that conditions of the model application period will be similar to those of the calibration period (Jiang et al., 2007; Xu, 1999b; Xu et al., 2005). However, altered rainfall-runoff relationship caused by climate and land use changes has also created some limitations and challenges in the use of hydrological models, which may cause the established models to become less skillful or lose their prediction ability in the new environment (Klemes, 1986). Therefore, it is essential to study the transferability of hydrological models in a changing environment.

Many studies on testing the applicability of hydrological models in changing climatic conditions have shown that many models do not have good temporal transferability, especially under non-stationary climatic conditions (Boorman and Sefton, 1997; Cornelissen et al., 2013; Eregno et al., 2013; Jiang et al., 2007; Li et al., 2012; Merz et al., 2011; Panagoulia and Dimou, 1997a, b). Moreover, the studies revealed that different hydrological models delivered different results when simulating hydrological responses to future climate change scenarios. Coron et al. (2012) used three lumped models (GR4J, MORDOR6 and SIMHYD) to simulate runoff processes in 216 watersheds in south-eastern Australia and found that the greater the climate difference between the calibration and validation periods, the worse was the transferability of the models. Broderick et al. (2016) used six lumped hydrological models to conduct a cross-validation study by dividing dry and wet years in the 37 watersheds of Ireland; results showed that model transferability depended on the selected catchment, tested scenarios and evaluation criteria. Oni et al. (2016) used historical wet and dry years as a proxy for expected future extreme conditions in a boreal catchment, demonstrating that runoff may be underestimated by at least 35% when model parameters were transferred from dry to wet years.

Hydrological models' spatial transferability has been studied using regionalisation methods (Bao et al., 2012; Merz and Bloschl, 2004; Parajka et al., 2013; Samuel et al., 2011; Swain and Patra, 2017; Yang et al., 2017, 2019, 2020). Yang et al. (2019) applied a lumped conceptual hydrological model (WASMOD) to investigate the transferability of regionalisation methods under changing climate conditions, based on 108 catchments in Norway. Lute and Luce (2017) built snow models of varying complexity in the western U.S. to evaluate model transferability in new locations and periods, indicating that the transferred models performed well in the new location with conditions similar to the trained location. They also found that simple to moderately complex models performed better than complex models when transferred to new locations in their study. Different results are reported by Yang et al. (2020) who tested spatial transferability of five conceptual hydrological models with varying number of parameters from 6 to 17, and concluded that the model with more parameters produced better results in most cases. A comprehensive survey of literature shows that there is no consistent conclusion about which regionalisation method or model performs best. Moreover, climate conditions are changing or are becoming non-stationary (IPCC, 2014), and under non-stationary climate conditions, the reliability of the model’s spatial transferability needs to be investigated. Therefore, it is very meaningful to further jointly study temporal and spatial transferabilities of different hydrological models under different climatic periods and in different basins.

The problem of general model transferability (spatial and temporal) has been recognised early as the major aim and the most difficult aspect of hydrological modelling (Klemes, 1986; Xu, 1999b). Despite this fact, less attention has been paid to the testing of this most important aspect, compared with many other modelling issues like manual versus automatic calibration, optimisation, regionalisation, etc. (Klemes, 1986; Xu, 1999b). In other words, operational testing of the models is not given the priority it deserves. Xu (1999b) made a preliminary attempt to evaluate temporal and spatial transferabilities of a lumped model in different simulation strategies; however, the study was limited by the number of models and data available at that time.

Above discussion reveals that although previous studies have explored transferability of hydrological models, some key issues are yet to be studied, which motivated the current study: (1) How do temporal and spatial transferabilities of hydrological models differ with the model complexity? (2) How do temporal and spatial transferabilities of hydrological models depend on different climates and underlying surface conditions of the basin? (3) What are the performance differences when the models are calibrated under dry/wet condition and transferred to wet/dry condition? To achieve these goals, five lumped hydrological models, including XAJ (Zhao, 1980), HBV (Bergstrom, 1976), SIMHYD (Chiew, 2002), IHACRES (Jakeman et al., 1990), and GR4J models (Perrin et al., 2003) with different complexities and flow generation methods are applied to two catchments in central-south China in this study. The temporal and spatial transferabilities of the five conceptual models are compared and analysed by using the split-sample, differential split-sample, proxy-basin and differential proxy-basin tests under stationary and changing conditions, including different climatic periods, different basins and their combinations. The rest of this paper is organised as follows. Section 2 introduces the study area and data. Section 3 provides the details about the five lumped models, and model calibration and validation methods. Section 4 presents and discusses the results corresponding to different simulation strategies. Finally, section 5 draws major conclusions and presents the limitations and possible future development of this study.

2. Study area and data

The study area for such a study must meet three requirements: (1) availability of long-term observation data; (2) extreme and variable climatic conditions to make it possible to select contrasting periods to test the capability of hydrological models under extreme conditions; and (3) significant differences of the underlying surface between the two basins. According to the requirements, Daxitan and Xiangxiang basins are selected as the study areas, whose location is shown in Fig. 1 and characteristics are listed in Table 1. The two basins are located in central-south China and cover a total area of 3010 km^2 and 5970 km^2, respectively. They both belong to a humid climate zone, which is also a necessary condition, as in practice, one will not expect to transfer a calibrated model in a humid region to an arid region, and vice versa. Affected by the monsoon climate and terrain, more than 65% of rainfall occurs in the rainy season from April to September for both basins. The mean annual rainfall, evapotranspiration and air temperature are 1560 mm, 847 mm, and 17°C, respectively, in Daxitan basin, and 1363 mm, 750 mm, 16.5°C, respectively, in Xiangxiang basin.

<Figure 1 here please >

< Table 1 here please >

Although both basins belong to the same climate zone, differences in their underlying surface, i.e., land covers and slope of the terrain, etc., are significant, as detailed in Table 1. Underlying surface conditions of Daxitan basin are more favourable for runoff generation and concentration than that of Xiangxiang basin, which can be verified by their runoff coefficients.

In this study, daily values of rainfall, pan evaporation, runoff and mean air temperature of the two basins for the period 1966–2015 are used to calibrate and validate the models. The daily mean air temperature is obtained from the National Meteorological Information Centre (http://data.cma.cn/), and other data are obtained from the published Yearly Hydrological Books of China. Considering the uneven distribution of meteorological stations, the Thiessen polygon method is used to calculate mean areal rainfall, evaporation and temperature of both basins as model input. These hydro-meteorological data are quality controlled by the Hydrology and Water Resources Bureau of Hunan Province, China and have been used in many other studies (e.g., Li et al., 2015; Xu et al., 2015a; Zeng et al., 2018).

<Figure 2 here please >

Figure 2 shows standardised annual rainfall and runoff (defined as deviation from the mean divided by the mean values) and standardised mean daily temperature (defined as deviation from the mean) and their five-year sliding results. Consistent changes between runoff and rainfall series can be seen, indicating that runoff is mainly driven by rainfall in the region. Temperature difference between the two basins is very small as they belong to the same climatic zone. The annual rainfall and runoff show no obvious trend but with distinct dry and wet periods, while the temperature of both basins showed a major upward trend over the entire record period, indicating that the selected period of 1966–2015 can be taken as the climate warming period to study the transferability of the hydrological models.

3. Hydrological models and methods

3.1. Hydrological models

Five conceptual hydrological models (XAJ, HBV, SIMHYD, IHACRES and GR4J), running at a daily time step, used to investigate transferability under changing conditions, are listed in Table 2. They are selected based on consideration of three aspects. First of all, the models are popular and commonly used in previous studies. Secondly, there are remarkable differences in their parameters and structures. Thus, they provide a good range of conceptual models available. Thirdly, as conceptual hydrological models are most widely used in assessing the impact on water resources in a changing environment, it is important to compare transferabilities between different conceptual hydrological models in changing environments (Broderick et al., 2016; Coron et al., 2012; Dakhlaoui et al., 2017; Fowler et al., 2016; Li et al., 2015, 2019; Vaze et al., 2010; Yang et al., 2020).

<Table 2 here please >

The XAJ model proposed by Zhao (1980) has been widely applied in humid and sub-humid regions (Jie et al., 2016; Lin et al., 2014; Yao et al., 2014; Zeng et al., 2016). In this model, hydrological processes can be divided into four groups: evapotranspiration, runoff production, separation of runoff components, and flow routing, linked to 15 parameters (Zhao, 1992). The HBV, originally developed by Swedish Meteorological and Hydrological Institute (SMHI) (Bergstrom, 1976), has been applied in many countries. The HBV model consists of a soil moisture routine, a response routine with three linear reservoir equations and a routing routine using the unit hydrograph (Osuch et al., 2019; Seibert, 1999). The SIMHYD model has nine parameters and includes three storages for interception loss, soil moisture and groundwater and the routing process (Chiew, 2002; Li et al., 2013). It considers different runoff production mechanisms for application in dry and wet areas. The IHACRES model is a lumped conceptual model based on the principle of unit hydrograph (Jakeman et al., 1990). It applies a transfer function/unit hydrograph approach to transform total rainfall to total runoff in two stages. In the first, a non-linear module is used to calculate effective rainfall by deducting the loss of rainfall, and then in the second linear module, effective rainfall is transformed into total runoff by fast and slow flows. The GR4J model is a simple lumped conceptual hydrological model with four parameters (Perrin et al., 2003). It routes runoff through a production reservoir, two linear unit hydrographs and a non-linear routing reservoir (Wang et al., 2018). Based on the difference in the routing time, the total runoff generation is divided into two runoff components according to the ratio of 9:1 (Perrin et al., 2003).

The five models are different in the way they conceptualise the hydrological processes and in their complexity (4–15 free parameters). The physical process is described in more detail and physical mechanism is more complex in XAJ, HBV, and SIMHYD models. The IHACRES model is a hybrid conceptual metric model, while GR4J is more simplified and empirical. The main feature of the runoff generation of XAJ and HBV models is that runoff is not generated until the soil moisture content of the aeration zone reaches its field capacity (i.e., saturation excess flow mechanism), while for SIMHYD model, surface runoff is not produced until the effective rainfall intensity is greater than the infiltration (i.e., infiltration excess mechanism). For the simulation of evaporation, XAJ model uses a three-layer evaporation model, while HBV and SIMHYD models use a one-layer model. Additionally, XAJ and HBV models consider uneven distribution of rainfall, but SIMHYD model does not. While

IHACRES model is designed to utilise the simplicity of the metric model to reduce the uncertainty of the hydrological model, it attempts to represent more detail of internal processes than is typical for a metric model (Coron et al., 2012). The GR4J model has the simplest structure between the models.

3.2. Validation methods for hydrological models

The test framework proposed by Klemes (1986) is used in this study. It is a typical test procedure based on selecting a specific contrast period from a long historical record to test a model's capability under changing conditions. The purpose of the test is to provide a set of basic safeguards and prevent the application of the model for tasks beyond its ability. The proposed scheme is called hierarchical because the modelling tasks are ordered according to their increasing complexity, and the demands of the test increase in the same direction. The four major categories are shown in Fig. 3.

<Figure 3 here please >

The split-sample test is the most common and fundamental operation to test model performance under stationary conditions. Available data are split into two parts; one for calibration and other for validation. Depending on the length of available sequences, segmentation can usually be done in a ratio of 1:1 or 7:3 (Klemes, 1986).

Proxy-basin test should be applied as a basic test when models are to be transferred between different basins, i.e., from a gauged to an ungauged catchment. The test needs to select at least two gauged basins in an adjacent region. The model is calibrated on a gauged basin and validated on the other gauged basin and vice versa. Only if the validation results of two basins are acceptable, the model might be used in the ungauged basin.

Differential split-sample test is used when a model is to be applied to simulate hydrological process under climate change in a gauged basin (Daggupati et al., 2015; Dakhlaoui et al., 2017; Fowler et al., 2016; Patil and Stieglitz, 2015; Westra et al., 2014; Zheng et al., 2018). This test is meaningful whenever a model is used to simulate runoff under conditions different from those corresponding to the available historical record. The main distinction from the split-sample test is that historical records are divided according to contrasting conditions of rainfall or other climatic variables, attempting to show that the model has general validity when used under climate change. For example, if increase in rainfall/temperature is the main change scenario in future, a dry/cold segment is selected to calibrate the model and wet/hot segment to validate it. The model with better validation results means better transferability under climate change.

The proxy-basin differential split-sample test is the most complicated test in Klemes' hierarchy. The model parameters need to be transferred under different climatic and spatial conditions. Such extensive transferability can be used as the ultimate objective and evaluation criterion of hydrological models. The specific test procedure is the combination of the proxy-basin and differential split-sample tests. First, two gauged basins A and B need to be selected, belonging to the same climate zone. Then, if increase in rainfall/temperature is the main change scenario in future, a dry/cold segment of basin A (B) is selected to calibrate the model and wet/hot segment of basin B (A) to validate it. The model with the best validation results will become the candidate model.

3.3. Model calibration and evaluation method

The shuffled complex evolution (SCE-UA; see Duan et al., 1992) algorithm, an effective global optimisation algorithm, is used to calibrate the models in this study. The algorithm is mainly based on the concept of information-sharing and natural biological evolution (Duan et al., 1994). It integrates the advantages of global sampling and complex evolution (Nelder and Mead, 1965). These characteristics can ensure the full use of sample information and greatly improve the convergence efficiency of the algorithm (Jeon et al., 2014). Therefore, it is widely used to calibrate parameters of conceptual hydrological models (Jie et al., 2018; Zeng et al., 2018).

In general, model parameters need to be calibrated with the criterion of making the difference between the simulated and observed runoff values from the historical record as small as possible. In this study, the objective function is a weighted combination of Nash efficiency coefficient (NS) and relative volume error (RE) proposed by Viney et al. (2009):

$$\mathrm{F} = \mathrm{NS} - 5 \times |\ln(1 + RE)|^{2.5} \tag{1}$$

where, NS and RE are shown in Eqs. (2) and (3), respectively. The optimal value of F is 1. This objective function is selected considering that it can effectively minimise RE, while at the same time maximise NS (Vaze et al., 2010).

$$\mathrm{NS} = 1 - \frac{\sum\left(Q_{obs}^{t} - Q_{sim}^{t}\right)^{2}}{\sum(Q_{obs}^{t} - \overline{Q_{obs}})^{2}} \tag{2}$$

$$\mathrm{RE} = \frac{\sum(Q_{sim}^{t} - Q_{obs}^{t})}{\sum Q_{obs}^{t}} \times 100\% \tag{3}$$

Here, Q_{obs}^{t} and Q_{sim}^{t} are the daily observed and simulated runoffs at time t, respectively, and $\overline{Q_{obs}}$ is the mean value of daily observed runoff. The NS represents the ratio between residual variance and observed data variance (Nash and Sutcliffe, 1970). To minimise the influence of initial condition on model performance, one year before the calibration period is used as the warm-up period.

The NS and RE are generally used to evaluate the accuracy of runoff simulation. Moriasi et al. (2007) proposed an NS and RE evaluation-grading category (Table 3) for evaluating model performance, widely used in runoff simulation in the world (Dakhlalla and Parajuli, 2016; Yang et al., 2019). However, numerical values of NS and RE are very different depending on, among others, geographic regions and hydrological models. For example, different threshold values of NS and RE are recommended in China for Hydrological Information and Hydrological Forecasting (HIHF) (Ministry of Water Resources, 2008). They are also listed in Table 3.

<Table 3 here please >

In order to have an objective criterion for evaluating the performance of transferability of hydrological models, we defined a new evaluation method based on the changes of NS and RE, shown in Table 4 and described as follows.

<Table 4 here please >

(1) If NS of the target catchment $NS_T \geqslant 0.70$ and RE of the target catchment $RE_T \leqslant 10\%$, the model is considered to have transferability regardless of the change range of NS (ΔNS) and RE ($|\Delta RE|$) between the calibrated and transferred models.

(2) If $NS_T < 0.70$ and $RE_T \leqslant 10\%$, the model is considered to have transferability when $\Delta NS \leqslant 0.2$; otherwise, it is considered to not have transferability.

(3) If $NS_T \geqslant 0.70$ and $RE_T > 10\%$, the model is considered to have transferability when $|\Delta RE| \leqslant 20\%$; otherwise, it is considered to not have transferability.

(4) If $NS_T < 0.70$ and $RE_T > 10\%$, the model is considered to have transferability when $\Delta NS \leqslant 0.2$

and |ΔRE| ≤ 20%; otherwise, it is considered to not have transferability.

4. Results and discussions

4.1. Spatial-temporal transferability tested by using odd and even years split-sample and proxy-basin methods

4.1.1Temporal transferability tested by using odd and even years split-sample test method

The split-sample test is carried out under stationary climate and basin conditions. In this experiment, the complete 50-year record is divided into odd and even years to avoid the influence caused by climate change. The Mann-Kendall (MK) test results reveal that the odd and even years runoff series in both basins do not have significant changing trends at 5% significance level and are considered to be stationary series. The models are calibrated using data of odd (even) years, and optimised parameters are used to simulate the runoff of even (odd) years.

The NS and RE values for different calibration and validation periods for the split-sample test are shown in Table 5. The five models perform similarly well for all calibrations, with all NS values greater than 0.79 and all RE values seem to be 0. All validations are slightly poorer but also show very good performance with all NS values exceeding 0.79 and RE values within ±10%. According to the proposed evaluation method in Table 4, performances of the five transferred models for this test are considered to be acceptable, as all NS > 0.70 and RE < 10%, indicating that the five models have temporal transferability under stationary conditions. Additionally, the difference between the results of different models is small.

<Table 5 here please >

4.1.2 **Spatial transferability tested by using proxy-basin method**

Similar to the split-sample test, the odd and even years described in Section 4.1.1 are used to obtain stationary climate conditions in this test. In this section, the spatial transferability test includes the proxy-basin and differential split-sample proxy-basin tests as shown in Figure 3, with the following combination scenarios: (1) Proxy-basin: calibrated on odd (even) years in Daxitan basin (A) and tested on odd (even) years in Xiangxiang basin (B) (A_{odd}-B_{odd} or A_{even}-B_{even}), calibrated on odd (even) years in Xiangxiang basin (B) and tested on odd (even) years in Daxitan basin (A) (B_{odd}-A_{odd} or B_{even}-A_{even}). (2) Differential split-sample proxy-basin: calibrated on odd (even) years in Daxitan basin (A) and tested on even (odd) years in Xiangxiang basin (B) (A_{odd}-B_{even} or A_{even}-B_{odd}), calibrated on odd (even) years in Xiangxiang basin (B) and tested on even (odd) years in Daxitan basin (A) (B_{odd}-A_{even} or B_{even}-A_{odd}).

Showing NS and RE values of different scenarios for the proxy-basin test, Fig. 4 reveals: (1) In most cases there is a slight increase in NS values when calibrated on Xiangxiang basin (B) and transferred to Daxitan basin (A), which include all four scenarios and almost all models (i.e., B_{odd}-A_{even}, B_{odd}-A_{odd}, B_{even}-A_{odd} and B_{even}-A_{even}), except B_{even}-A_{odd} for HBV model. This slight increase in NS values is because the calibration result of Daxitan basin as measured by NS is about 0.05–0.1 higher than that of Xiangxiang basin, as seen in Table 5. In this case, the XAJ model performed best for all the scenarios and GR4J is the worst for two scenarios (B_{even}-A_{odd} and B_{even}-A_{even}). (2) In all scenarios, there is a big drop of NS values when calibrated on Daxitan basin (A) and transferred to Xiangxiang basin (B) (i.e., A_{odd}-B_{even}, A_{odd}-B_{odd}, A_{even}-B_{odd} and A_{even}-B_{even}). In this case, the XAJ model performed best and GR4J is the worst for all scenarios. (3) In terms of RE values, there is a 10% to 20% negative bias when the models are calibrated on Xiangxiang basin and transferred to Daxitan basin, which is true for all four scenarios and five models (i.e., B_{odd}-A_{even}, B_{odd}-A_{odd}, B_{even}-A_{odd} and B_{even}-A_{even}). The opposite is true when the models are calibrated on Daxitan basin and transferred to Xiangxiang basin, where there is 10% to 30% positive bias depending on the model (i.e., A_{odd}-B_{even}, A_{odd}-B_{odd}, A_{even}-B_{odd} and A_{even}-B_{even}). (4) According to the evaluation criterion defined in Table 4, transferability of all five models is not accepted under A_{odd}-B_{even} scenario; GR4J does not have transferability under A_{even}-B_{odd} scenario, as its RE > 10% and |ΔRE| > 20%. The GR4J and SIMHYD models do not have transferability under A_{odd}-B_{odd} scenario as their RE > 10% and |ΔRE| > 20%; only IHACRES shows transferability under A_{even}-B_{even} scenario, as its NS > 0.7 and |ΔRE| < 20%. Only performances of transferred GR4J and SIMHYD are not acceptable under scenario B_{even}-A_{odd} because their RE > 10% and |ΔRE| > 20%.

<Figure 4 here please >

Above discussion reveals that performances of the five models in Daxitan basin (A), with a runoff coefficient of 0.58, are consistently and significantly better than those in Xiangxiang basin (B) with a runoff coefficient of 0.44, in the calibration period. This is an important reason behind the sharp drop in NS values and a positive bias when the models are calibrated on Daxitan basin (A) and transferred to Xiangxiang basin (B) (i.e., A_{odd}-B_{even}, A_{odd}-B_{odd}, A_{even}-B_{odd} and A_{even}-B_{even}). On the contrary, there is a negative bias when the models are calibrated on Xiangxiang basin (B) with lower runoff coefficient and transferred to Daxitan basin (A) with a higher runoff coefficient (i.e., B_{odd}-A_{even}, B_{odd}-A_{odd}, B_{even}-A_{odd} and B_{even}-A_{even}). In this case, there is even a slight increase in NS values; however, transferred NS values in Daxitan basin (A) are still lower than calibrated values in the basin (Table 5). These results mean that when a model is transferred from a basin with favorable runoff generation conditions to one with less favorable runoff generation conditions, a big drop in NS values may be expected.

4.2. Spatial-temporal transferability tested by using driest and wettest periods using split-sample and proxy-basin methods

4.2.1 Temporal transferability tested by using driest and wettest periods using split-sample method

This section verifies the prediction ability of the hydrological models in transferring from more contrasted periods of five consecutive driest (wettest) years to five consecutive wettest (driest) years, using the differential split-sample test. The consecutive driest and wettest five-year records from the 50-year historical dataset are selected for this test. As runoff generation is mainly driven by precipitation in both basins, the driest and wettest hydrological periods are chosen according to the sum of consecutive five-year annual rainfall amounts from the rainfall series. Table 6 shows the mean monthly rainfall, runoff, temperature and runoff coefficient of the selected consecutive five-year driest and wettest periods. Compared with the driest hydrological period, the rainfall of the wettest period increases by nearly 20% and the runoff increases by more than 50%.

<Table 6 here please >

To perform this differential split-sample test, the driest and wettest periods are in turn taken as calibration and transfer periods in the study basins, whose results are shown in Fig. 5. Figure 5 reveals that results of transferred models in Daxitan (A) and Xiangxiang basins (B) are quite different. In Daxitan basin (A), the five transferred models show slight decrease in NS values, where GR4J model has the biggest drop in A_{wet}-A_{dry} scenario, but the transferred NS value is still higher than 0.77. As for the RE value, both positive and negative biases are seen depending on the model and the scenario. However, only in the A_{wet}-A_{dry} scenario, the positive bias of GR4J model exceeded the threshold limit of 20%, and all other models and scenarios are considered as acceptable. In Xiangxiang basin (B), when the driest period is used as the calibration period, performances of the five models when transferred to the wettest period are satisfactory, as their NS > 0.7 and $|\Delta RE| < 20\%$. When the five models are transferred from the wettest to the driest period, XAJ, IHACRES and GR4J have temporal transferability, while HBV and SIMHYD do not have temporal transferability as their $|\Delta RE| > 20\%$.

<Figure 5 here please >

From these results, the models perform better in Daxitan basin (A) than they do in Xiangxiang basin (B). It also can be found that they perform better when transferred from the driest to the wettest period in both basins than when transferred from the wettest to the driest period.

4.2.2 Spatial transferability in contrast periods by using proxy-basin test

In order to further investigate spatial transferability of the models under contrast hydrological conditions, a process similar to Section 4.1.2 is performed here, except that the calibration and transfer periods are replaced by the consecutive wettest and driest five-year periods, whose results are shown in Fig. 6.

<Figure 6 here please >

Figure 6 reveals that the five transferred models perform well under B_{wet}-A_{dry} and A_{dry}-B_{wet} scenarios, as all NS values of the transferred models > 0.7 and $|\Delta RE| < 20\%$. Under A_{wet}-B_{dry}, B_{dry}-A_{wet} and

B_{dry}-A_{dry} scenarios, the transferability of XAJ and HBV is acceptable, but for other three models it is not acceptable, as their |ΔRE| > 20% in the transferred models. The five models perform poorest under A_{dry}-B_{dry} scenario, as all ΔNS > 0.2 and |ΔRE| > 20% in the transferred models. Under B_{wet}-A_{wet} and A_{wet}-B_{wet} scenarios, GR4J and IHACRES do not have transferability as their |ΔRE| > 20%.

It can be concluded from the results in Fig. 6 that all five models are verified to have transferability under scenarios B_{wet}-A_{dry} and A_{dry}-B_{wet}. While SIMHYD, IHRCRES and GR4J do not have transferability, XAJ and HBV have transferability under other three scenarios: B_{dry}-A_{wet}, A_{wet}-B_{dry} and B_{dry}-A_{dry}. Transferred results of SIMHYD, IHRCRES and GR4J deteriorate sharply under A_{wet}-B_{dry} scenario, as seen from their large ΔNS and |ΔRE| values. Under A_{dry}-B_{dry} scenario, all five models lose their simulation ability as reflected by low NS values and high RE values of the transferred models.

4.3. Transferability test under the most extreme conditions

In previous sections, studies on transferability of the hydrological model are carried out under long-term (Section 4.1) and contrast consecutive five-year wettest and driest periods (Section 4.2). As the occurrence of extreme climatic or hydrological events has been on the rise in recent years (Groisman et al., 2004; Westra et al., 2013), it is of great significance to study the mechanism and influence of extreme hydrological events on an annual scale. Here, the driest and wettest years of the 50-year series will be selected to verify the ability of hydrological models to simulate extreme hydrological events on an annual scale, in order to answer the question: Do the models still have the similar predictive ability when calibrated under the driest or wettest year condition? To answer this question, we design and perform another transferability experiment under driest and wettest year of the

50-year series. Characteristics of the driest and wettest years are shown in Table 7. Climate difference is more significant than the consecutive five-year record, as expected (shown in Table 6). Rainfall and temperature variations of the two basins have exceeded 60% and 1.0°C between the driest and wettest years, respectively. Compared with the driest year, runoff changes in the wettest year exceed 160% and 290% for Daxitan (A) and Xiangxiang basins (B), respectively. They can represent the main characteristics of annual extreme hydrological events in both basins.

<Table 7 here please >

The results of this experiment obviously magnify the runoff simulation error shown in Section 4.2 (comparing Figs. 6 and 7). Figure 7 reveals that for temporal transferability (up panel), the five models do not have transferability under A_{wet}-A_{dry} and B_{wet}-B_{dry} scenarios as their $\Delta NS > 0.2$ and $|\Delta RE| > 20\%$; only SIMHYD does not have transferability under A_{dry}-A_{wet} scenario as its $|\Delta RE| > 20\%$. The SIMHYD and IHACRES models do not have transferability under B_{dry}-B_{wet} scenario as their $|\Delta RE| > 20\%$, although their NS values are higher than 0.70. For spatial transferability (lower panel), five transferred models perform well under B_{wet}-A_{wet} and A_{wet}-B_{wet} scenarios because their $NS > 0.70$ and $|\Delta RE| < 20\%$. For B_{dry}-A_{dry} scenario, only HBV and XAJ perform well as their $NS > 0.70$ and $|\Delta RE| < 10\%$. For A_{dry}- B_{dry} scenario, HBV, XAJ and SIMHYD perform well as their $NS > 0.70$ and $|\Delta RE| < 20\%$. For temporal and spatial transferabilities (middle panel), most models perform poorly and do not have transferability because of large change in their NS and RE values, especially under A_{wet}-B_{dry} scenario. The HBV performs well when transferred from the driest year in Xiangxiang Basin (B) to the wettest year in Daxitan basin (A). The GR4J performs well under B_{wet}-A_{dry} and A_{dry}-B_{wet} scenarios, as its $NS > 0.70$ and $|\Delta RE| < 20\%$. The XAJ and HBV also show transferability under A_{dry}-B_{wet}

scenarios, as its NS > 0.70 and $|\Delta RE| < 20\%$. It is concluded that all transferred models show greater uncertainties in different scenarios under yearly extreme scenarios, especially from the wettest to the driest year. According to the ranges of changes in NS and RE values, the five models perform worse when transferred from the wettest to the driest year, although their calibration performances are very good in both the driest and wettest years.

<Figure 7 here please >

In order to further test the transferability of the models under the driest and wettest years, five typical years with quantiles of 5%, 25%, 50%, 75% and 95% are selected from the 50 annual runoff series, sorted from low to high. The five models are transferred from the driest or wettest year to the five typical years between the two basins, which will generate 200 cases for this test. Based on the proposed evaluation method, the case that the transferability of one model is accepted is recorded as 1, otherwise 0, counted for each model and listed in Table 8. For temporal transferability, the five models show better performances in Daxitan basin (A) than in Xiangxiang basin (B), consistent with the fact that Daxitan basin (A) has a favourable runoff generation condition. For spatial transferability, similar results can be found wherein the five models perform poorly when transferred from Daxitan basin (A) to Xiangxiang basin (B); only XAJ and HBV perform well when transferred from the driest year in Xiangxiang basin (B) to five typical years in Daxitan basin (A). Other three models show no transferability in these scenarios. Five models except IHACRES show good performance when transferred from the wettest year in Xiangxiang basin (B) to the five typical years in Daxitan basin (A). According to the results given in Table 8, the transferred XAJ model performs the best, while IHACRES is the worst between the five models. The main reason is that IHCRES in the calibration driest or wettest year has a lower NS value than other models in both basins, according to Fig. 7.

<Table 8 here please>

4.4 Comprehensive evaluation of spatial and temporal transferabilities of the hydrological models

According to the results from Sections 4.1 to 4.3, in total, 76 change scenarios are used to compare spatial and temporal transferability differences of the five models. From the above discussion it is seen that the model evaluation method defined in this study (Table 4) has been proved to be useful since it simultaneously evaluates the NS and RE values together with the changes in them. In order to synthetically compare the transferability of the five models based on the results from Sections 4.1 to 4.3, the 76 scenarios are divided into different categories as shown in Table 9. Results of the transferred models with acceptable transferability are recorded as 1, while results with unacceptable transferability as 0, counted for each scenario and listed in Table 9.

<Table 9 here please >

For temporal transferability, XAJ, GR4J and HBV show good transferability as their recorded numbers are more than 22 of the total 32 scenarios. For spatial transferability, there is a big difference between the complex and simple models. For example, there are 10 scenarios for XAJ and HBV, while only four for GR4J to have acceptable transferability. Similar results can be found in temporal-spatial categories. This is helpful in selecting models to transfer runoff in spatial and temporal dimensions. In this study, the selected two basins are adjacent and belong to the same climatic zone; thus, their precipitation regimes can be regarded as similar, while, their terrain and land covers are significantly different, as detailed in Table 1. The percentage of the model cases recorded with acceptable transferability from B to A (59.1%) is much higher than 46.4% of the models transferred from A to B, meaning that the underlying surface conditions of the target basin are important impactors to the spatial transfer of the hydrological models. Results of the five transferred models between the driest and wettest conditions are also quite different. There are 70% of the model cases with acceptable transferability from the driest to the wettest period, while the number is 37.5% from the wettest to the driest period according to the data in Table 9. As there are only two basins selected to compare the models' transferability in this study, the results and findings need to be further verified by selecting more basins.

5. Summary and conclusions

Hydrological models have been widely used in hydrology and water resources management. It is also the most important tool for hydrologic prediction in ungauged basins, and for studying the impact of climate change and human activities on hydrology. However, models have different conceptualisation schemes and mathematical representation of the hydrologic processe, which determines that the prediction ability of each model is different. When a model is transferred to another basin or period, the question to be answered is whether the model still has the same ability of simulation and prediction as in the calibration period? To answer it, exploratory research on temporal and spatial transferabilities of hydrological models is carried out in this study. To achieve the goal, five hydrological models with different complexities of structure are used to illustrate differences in transferability between the models. Two basins with different underlying surface conditions are adopted to set up 76 transferability scenarios, including odd and even stationary series, driest and wettest series, which can reflect temporal and spatial changes between the two basins. Simulation results of these scenarios are evaluated by a new evaluation method proposed in this study, and the main conclusions are as follows.

(1) The proposed evaluation method for transferability based on absolute and relative changes of NS and RE values is used to judge whether the transferred model has the ability of simulation and prediction. The study proves the proposed evaluation method is effective in evaluating the transfer ability of the model, as it provides an objective and quantitative measure.

(2) For temporal transferability under stationary condition, all five models show good performances. But for spatial transferability, the complex models (XAJ and HBV) have much better performances than the simple model (GR4J) in this study, as there are 10 of the total 12 scenarios for XAJ and HBV, while only four for GR4J to have acceptable transferability.

(3) The difference in underlying surface conditions in the target basin affects the spatial transferability of the models in such a way that better results are obtained when the model is transferred to a basin with favourable runoff generation condition than the opposite case.

(4) In the transfer between the driest and wettest periods, the error is larger when the models are calibrated in the wettest period and transferred to the driest period than the opposite approach. This provides good reference information for the study on the impact of climate change on hydrological extremes.

There are, however, also some limitations in this study, which warrant further study. For example, the uncertainties of the models are not considered in this study. In the study of spatial transferability, only two adjacent basins are considered, and the conclusions obtained from them may not be generalized until more basins are evaluated. This study only considers different types of lumped models, however, distributed models are also widely used in changing environments. Therefore, a comprehensive comparison of the differences between distributed and lumped models in spatial and temporal transferabilities will be helpful to enrich the research results of the spatial and temporal transferabilities of hydrological modelling in changing environments.

Acknowledgements

The study was financially supported by the National Natural Science Fund of China (51539009) and National Key Research and Development Program (2017YFA0603702), and the Research Council of Norway (FRINATEK Project 274310).

References

Alcamo, J., Floerke, M., Maerker, M., 2007. Future long-term changes in global water resources driven by socio-economic and climatic changes. Hydrol. Sci. J. 52 (2), 247-275.

Arnell, N.W., 2004. Climate change and global water resources: SRES emissions and socio-economic scenarios. Global Environ. Chang. 14 (1), 31-52.

Bao, Z., Zhang, J., Liu, J., Fu, G., Wang, G., He, R., Yan, X., Jin, J., Liu, H., 2012. Comparison of regionalization approaches based on regression and similarity for predictions in ungauged catchments under multiple hydro-climatic conditions. J. Hydrol. 466, 37-46.

Bergstrom, S., 1976. Development and Application of a Conceptual Runoff Model for Scandinavian Catchments. SMHI Reports RHO (No.7, Norrkoping).

Boorman, D.B., Sefton, C., 1997. Recognising the uncertainty in the quantification of the effects of climate change on hydrological response. Clim. Change 35 (4), 415-434.

Broderick, C., Matthews, T., Wilby, R.L., Bastola, S., Murphy, C., 2016. Transferability of hydrological models and ensemble averaging methods between contrasting climatic periods. Water Resour. Res. 52 (10), 8343-8373.

Chen, L., Chang, J., Wang, Y., Zhu, Y., 2019. Assessing runoff sensitivities to precipitation and temperature changes under global climate-change scenarios. Hydrol. Res. 50 (1), 24-42.

Chen, Y., Xu C.Y., Chen, X., Xu, Y., Yin, Y., Gao, L., Liu, M., 2019. Uncertainty in simulation of land-use change impacts on catchment runoff with multi-timescales based on the comparison of the HSPF and SWAT models. J. Hydrol. 573, 486-500.

Chiew, F.H.S., 2002. Application and testing of the simple rainfall-runoff model SIMHYD.

Cornelissen, T., Diekkrueger, B., Giertz, S., 2013. A comparison of hydrological models for assessing the impact of land use and climate change on discharge in a tropical catchment. J. Hydrol. 498, 221-236.

Coron, L., Andreassian, V., Perrin, C., Lerat, J., Vaze, J., Bourqui, M., Hendrickx, F., 2012. Crash testing hydrological models in contrasted climate conditions: An experiment on 216 Australian catchments. Water Resour. Res. 48 (W05552).

Daggupati, P., Pai, N., Ale, S., Douglas-Mankin, K.R., Zeckoski, R.W., Jeong, J., Parajuli, P.B., Saraswat, D., Youssef, M.A., 2015. A recommended calibration and validation strategy for hydrologic and water quality models. T. Asabe 58 (6), 1705-1719.

Dakhlalla, A. O., Parajuli, P. B., 2016. Evaluation of the best management practices at the watershed scale to attenuate peak streamflow under climate change scenarios. Water Resour. Manag. 30 (3), 963-982.

Dakhlaoui, H., Ruelland, D., Tramblay, Y., Bargaoui, Z., 2017. Evaluating the robustness of conceptual rainfall-runoff models under climate variability in northern Tunisia. J. Hydrol. 550, 201-217.

Doll, P., 2002. Impact of climate change and variability on irrigation requirements: A global perspective. Clim. Change 54 (3), 269-293.

Duan, Q.Y., Sorooshian, S., Gupta, V.K., 1992. Effective and efficient global optimization for conceptual rainfall-runoff models. Water Resour. Res. 28 (4), 1015-1031.

Duan, Q.Y., Sorooshian, S., Gupta, V.K., 1994. Optimal use of the SCE-UA global optimization method for calibrating watershed models. J. Hydrol. 158 (3-4), 265-284.

Eregno, F.E., Xu, C.-Y., Kitterod, N.O., 2013. Modeling hydrological impacts of climate change in different climatic zones. International Journal of Climate Change Strategies and Management 5 (3), 344-365.

Fan, M., Mawuko, D.O., Shibata, H., Ou, W., 2019. Spatial conservation areas for water yield hydrological ecosystem services with their economic values effects under climate change: a case study of Teshio watershed located in northernmost of Japan. Hydrol. Res. 50 (6), 1679–1709.

Fowler, K.J.A., Peel, M.C., Western, A.W., Zhang, L., Peterson, T.J., 2016. Simulating runoff under changing climatic conditions: Revisiting an apparent deficiency of conceptual rainfall-runoff models. Water Resour. Res. 52 (3), 1820-1846.

Frich, P., Alexander, L.V., Della-Marta, P., Gleason, B., Haylock, M., Tank, A., Peterson, T., 2002. Observed coherent changes in climatic extremes during the second half of the twentieth century. Clim. Res. 19 (3), 193-212.

Gleick, P.H., 1986. Methods for evaluating the regional hydrologic impacts of global climatic changes. J. Hydrol. 88 (1-2), 97-116.

Groisman, P. Y., Knight, R. W., Easterling, D. R., Karl, T. R., Razuvaev, V. N., 2004. Trends in precipitation intensity in the climate record. J Climate 18 (9), 1326–1350.

Guo, Y.X., Fang, G.H., Wen, X., Lei, X.H., Yuan, Y., Fu, X.Y., 2019. Hydrological responses and adaptive potential of cascaded reservoirs under climate change in Yuan River Basin. Hydrol. Res. 50 (1), 358–378.

Jakeman, A.J., Littlewood, I.G., Whitehead, P.G., 1990. Computation of the instantaneous unit-hydrograph and identifiable component flows with application to 2 small upland catchments. J. Hydrol. 117 (1-4), 275-300.

Jeon, J., Park, C., Engel, B.A., 2014. Comparison of Performance between Genetic Algorithm and SCE-UA for Calibration of SCS-CN Surface Runoff Simulation. Water-Sui. 6 (11), 3433-3456.

Jiang, T., Chen, Y.D., Xu, C.-Y., Chen, X., Chen, X., Singh, V.P., 2007. Comparison of hydrological impacts of climate change simulated by six hydrological models in the Dongjiang Basin, South China. J. Hydrol. 336 (3-4), 316-333.

Jie, M.X., Chen, H., Xu, C.-Y., Zeng, Q., Tao, X.E., 2016. A comparative study of different objective functions to improve the flood forecasting accuracy. Hydrol. Res. 47(4), 718-735.

Jie, M., Chen, H., Xu, C.-Y., Zeng, Q., Chen, J., Kim, J.S., Guo, S., Guo, F., 2018. Transferability of Conceptual Hydrological Models Across Temporal Resolutions: Approach and Application. Water Resour. Manag. 32 (4), 1367-1381.

Klemes, V., 1986. Operational testing of hydrological simulation models. Hydrol. Sci. J. 31 (1), 13-24.

Li, C.Z., Zhang, L., Wang, H., Zhang, Y.Q., Yu, F.L., Yan, D.H., 2012. The transferability of hydrological models under nonstationary climatic conditions. Hydrol. Earth Syst. Sc. 16 (4), 1239-1254.

Li, F., Zhang, Y., Xu, Z., Teng, J., Liu, X., Liu, W., Mpelasoka, F., 2013. The impact of climate change on runoff in the southeastern Tibetan Plateau. J. Hydrol. 505, 188-201.

Li, H., Beldring, S, Xu, C.-Y., 2015. Stability of model performance and parameter values on two catchments facing changes in climatic conditions. Hydrol. Sci. J. 60(7-8), 1317-1330.

Li, Y.Y., Luo, L.F., Wang, Y.M., Guo, A.J., Ma, F., 2019. Spatiotemporal impacts of land use land cover changes on hydrology from the mechanism perspective using SWAT model with time-varying parameters. Hydrol. Res. 50 (1), 244-261.

Lin, K.R., Lv, F., Chen, L., Singh, V.P., Zhang, Q., Chen, X., 2014. Xinanjiang model combined with Curve Number to simulate the effect of land use change on environmental flow. J. Hydrol. 519, 3142-3152.

Lu, W., Qin, X.S., 2020. Integrated framework for assessing climate change impact on extreme rainfall and the urban drainage system. Hydrol. Res. 51 (1), 77-89. https://doi.org/10.2166/nh.2019.233.

Lute, A.C., Luce, C.H., 2017. Are model transferability and complexity antithetical? Insights from validation of a variable-complexity empirical snow model in space and time. Water Resour. Res. 53, 8825-8850.

Ma, Z., Kang, S., Zhang, L., Tong, L., Su, X., 2008. Analysis of impacts of climate variability and human activity on streamflow for a river basin in arid region of northwest China. J. Hydrol. 352 (3-4), 239-249.

Merz, R., Bloschl, G., 2004. Regionalisation of catchment model parameters. J. Hydrol. 287 (1-4), 95-123.

Merz, R., Parajka, J., Bloeschl, G., 2011. Time stability of catchment model parameters: Implications for climate impact analyses. Water Resour. Res. 47 (W02531).

Milly, P.C.D., Betancourt, J., Falkenmark, M., Hirsch, R.M., Kundzewicz, Z.W., Lettenmaier, D.P., Stouffer, R.J., 2008. Stationarity Is Dead: Whither Water Management? Science 319 (5863), 573-574.

Ministry of Water Resources, China, 2008. Standard for hydrological information and hydrological forecasting. Beijing, China Water & Power Press. GB/T 22482-2008.

Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D., Veith, T.L., 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. T. Asabe 50 (3), 885-900.

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.

Nelder, J.A., Mead, R., 1965. A simplex-method for function minimization. Comput. J. 7 (4), 308-313.

Oni, S., Futter, M., Ledesma, J., Teutschbein, C., Buttle, J., Laudon, H., 2016. Using dry and wet year hydroclimatic extremes to guide future hydrologic projections. Hydrol. Earth Syst. Sc. 20 (7), 2811-2825.

Osuch, M., Wawrzyniak, T., Nawrot, A., 2019. Diagnosis of the hydrology of a small Arctic permafrost catchment using HBV conceptual rainfall-runoff model. Hydrol. Res. 50 (2), 459-478.

Panagoulia, D., Dimou, G., 1997a. Linking space-time scale in hydrological modelling with respect to global climate change .1. Models, model properties, and experimental design. J. Hydrol. 194 (1-4), 15-37.

Panagoulia, D., Dimou, G., 1997b. Linking space-time scale in hydrological modelling with respect to global climate change .2. Hydrological response for alternative climates. J. Hydrol. 194 (1-4), 38-63.

Parajka, J., Viglione, A., Rogger, M., Salinas, J.L., Sivapalan, M., Bloeschl, G., 2013. Comparative assessment of predictions in ungauged basins - Part 1: Runoff-hydrograph studies. Hydrol. Earth Syst. Sc. 17 (5), 1783-1795.

Patil, S.D., Stieglitz, M., 2015. Comparing Spatial and temporal transferability of hydrological model parameters. J. Hydrol. 525, 409-417.

Perrin, C., Michel, C., Andréassian, V., 2003. Improvement of a parsimonious model for streamflow simulation. J. Hydrol. 279 (1-4), 275-289.

Ragettli, S., Tong, X., Zhang, G., Wang, H., Zhang, P., Stähli, M., 2020. Climate change impacts on summer flood frequencies in two mountainous catchments in China and Switzerland. Hydrol. Res.

in press. https://doi.org/10.2166/nh.2019.118.

Samuel, J., Coulibaly, P., Metcalfe, R.A., 2011. Estimation of Continuous Streamflow in Ontario Ungauged Basins: Comparison of Regionalization Methods. J. Hydrol. Eng. 16 (5), 447-459.

Schulze, R.E., 1997. Impacts of global climate change in a hydrologically vulnerable region: Challenges to South African hydrologists. Prog. Phys. Geog. 21 (1), 113-136.

Seibert, J., 1999. Regionalisation of parameters for a conceptual rainfall-runoff model. Agr. Forest Meteorol. 98-99, 279-293.

Swain, J.B., Patra, K.C., 2017. Streamflow estimation in ungauged catchments using regionalization techniques. J. Hydrol. 554, 420-433.

Vaze, J., Post, D.A., Chiew, F.H.S., Perraud, J.M., Viney, N.R., Teng, J., 2010. Climate non-stationarity - Validity of calibrated rainfall-runoff models for use in climate change studies. J. Hydrol. 394 (3-4), 447-457.

Viney, N.R., Perraud, J., Vaze, J., Chiew, F.H.S., Post, D.A., Yang, A., 2009. The usefulness of bias constraints in model calibration for regionalization to ungauged catchments. In: Proc.18th World IMACS/MODSIM Congress. International Environmental Modelling and Software Society, Cairns, Australia, pp. 3421-3427.

Wang, H.M., Chen, J., Cannon, A.J., Xu, C.-Y., Chen, H., 2018. Transferability of climate simulation uncertainty to hydrological impacts. Hydrol. Earth Syst. Sc. 22, 3739-3759.

Westra, S., Alexander, L.V., Zwiers, F.W., 2013. Global increasing trends in annual maximum daily precipitation. J. Climate 26 (11), 3904-3918.

Westra, S., Thyer, M., Leonard, M., Kavetski, D., Lambert, M., 2014. A strategy for diagnosing and interpreting hydrological model nonstationarity. Water Resour. Res. 50 (6), 5090-5113.

Xiong, F., Guo, S., Liu, P., Xu, C.-Y., Zhong, Y., Yin, J., He, S., 2019. A general framework of design flood estimation for cascade reservoirs in operation period. J. Hydrol. 577 (UNSP 124003).

Xu, C.-Y., 1999b. Operational testing of a water balance model for predicting climate change impacts. Agr. Forest Meteorol. 98-9 (SI), 295-304.

Xu, C.-Y., Singh, V.P., 2004. Review on regional water resources assessment models under stationary and changing climate. Water Resour. Manag. 18 (6), 591-612.

Xu, C.-Y., Widen, E., Halldin, S., 2005. Modelling hydrological consequences of climate change - Progress and challenges. Adv. Atmos. Sci. 22 (6), 789-797.

Xu, H., Xu, C.-Y., Sælthun, N. R., Xu, Y., Zhou, B., Chen, H., 2015a. Entropy theory based multi-criteria resampling of rain gauge networks for hydrological modelling–A case study of humid area in southern China. J. Hydrol. 525, 138-151.

Yang, X., Magnusson, J., Rizzi, J., Xu, C.-Y., 2017. Runoff prediction in ungauged catchments in Norway: comparison of regionalization approaches. Hydrol. Res. 49 (2), 487-505.

Yang, X., Magnusson, J., Xu, C.-Y., 2019. Transferability of regionalization methods under changing climate. J. Hydrol. 568, 67-81.

Yang, X., Magnusson, J., Huang, S.C., Beldring, S., Xu, C.-Y., 2020. Dependence of regionalization methods on the complexity of hydrological models in multiple climatic regions. J. Hydrol. 582,124357.

Yao, C., Zhang, K., Yu, Z.B., Li, Z.J., Li, Q.L., 2014. Improving the flood prediction capability of the Xinanjiang model in ungauged nested catchments by coupling it with the geomorphologic instantaneous unit hydrograph. J. Hydrol. 517, 1035-1048.

Ye, X., Zhang, Q., Liu, J., Li, X., Xu, C.-Y., 2013. Distinguishing the relative impacts of climate change and human activities on variation of streamflow in the Poyang Lake catchment, China. J. Hydrol. 494, 83-95.

Zeng, Q., Chen, H., Xu, C.-Y., Jie, M.X., Hou, Y.K., 2016. Feasibility and uncertainty of using conceptual rainfall runoff models in design flood estimation. Hydrol. Res. 47(4), 701-717. doi: 10.2166/nh.2015.069.

Zeng, Q., Chen, H., Xu, C.-Y., Jie, M., Chen, J., Guo, S., Liu, J., 2018. The effect of rain gauge density and distribution on runoff simulation using a lumped hydrological modelling approach. J. Hydrol. 563, 106-122.

Zhang, Z., Chen, X., Xu, C.-Y., Yuan, L., Yong, B., Yan, S., 2011. Evaluating the non-stationary relationship between rainfall and streamflow in nine major basins of China during the past 50 years. J. Hydrol. 409 (1-2), 81-93.

Zhang, Z., Xu, C.-Y., El-Tahir, M.E., Cao, J., Singh, V.P., 2012. Spatial and temporal variation of rainfall in Sudan and their possible causes during 1948-2005. Stoch. Env. Res. Risk a. 26 (3), 429-441.

Zhao, R.J., 1992. The Xinanjiang model applied in China. J. Hydrol. 135, 371–381.

Zhao, R.J., Zhuang, Y.L., Fang, L.R., Liu, X.R., Zhang, Q.S., 1980. The Xinanjiang Model. In:

Hydrological Forecasting, IAHS Publication No. 129. IAHS Press, Wallingford: 351–356.

Zheng, F., Maier, H.R., Wu, W., Dandy, G.C., Gupta, H.V., Zhang, T., 2018. On Lack of Robustness in Hydrological Model Development Due to Absence of Guidelines for Selecting Calibration and Evaluation Data: Demonstration for Data-Driven Models. Water Resour. Res. 54 (2), 1013-1030.

Tables

Table 1. Characteristics of the climate, terrain and vegetation cover of the study basins.

Table 2. Structural components of the five lumped conceptual rainfall-runoff models.

Table 3. Classification of model performance into categories with limits following Moriasi et al. (2007) and Standard of HIHF(Ministry of Water Resources, China, 2008) for NS and RE.

Table 4. Evaluation criteria for hydrological model transferability based on Table 3.

Table 5. Comparison of the statistics results of the five models for temporal transferability test by using odd and even years split-sample test.

Table 6. The hydro-climatic variables with the driest and wettest consecutive 5-year periods.

Table 7. The hydro-climatic variables with the driest and wettest 1-year periods.

Table 8. The sensitivity analysis of the temporal and spatial transferability of the hydrological models under driest year and wettest year.

Table 9. The number of models with transferability under different scenarios.

Figures

Fig. 1. Location and underlying surface mapping of the study area.

Fig. 2. Standardized mean annual precipitation (P), air temperature (T), and mean annual runoff for Daxitan and Xiangxiang basins. The bar chart shows the results of each year, and the red line shows the results of every consecutive 5-year period by moving the window by one year.

Fig. 3. Hierarchical approach for operational testing of hydrological simulations.

Fig. 4. Comparison of the NS and RE values of the five models for the proxy-basin test using the odd and even years. (A = Daxitan; B = Xiangxiang; A_{odd}-B_{even} indicates the calibration in odd years at Daxitan and transfer in even years at Xiangxiang, etc.; Dotted line means △NS > 0.2 or |△RE| > 20%)

Fig. 5. Comparison of the NS and RE values of the five models for the differential split-sample test using the driest and wettest consecutive 5-year records. (A = Daxitan; B = Xiangxiang; A_{dry}-A_{wet} indicates the calibration in dry years at Daxitan and transfer in wet years at Daxitan, etc.; Dotted line means △NS > 0.2 or |△RE| > 20%)

Fig. 6. Comparison of the NS and RE values of the five models for proxy-basin test using the driest and wettest consecutive 5-year records. (A = Daxitan; B = Xiangxiang; A_{dry}-B_{wet} indicates the calibration in dry years at Daxitan and transfer in wet years at Xiangxiang, etc.; Dotted line means △NS > 0.2 or |△RE| > 20%)

Fig. 7. Comparison of the NS value and RE value of the five models using the driest and wettest 1-year records. (A = Daxitan; B = Xiangxiang; A_{dry}-B_{wet} indicates the calibration in dry years at Daxitan and transfer in wet years at Xiangxiang, etc.; Dotted line means △NS > 0.2 or |△RE| > 20%)

*Credit Author Statement

Wushuang Yang: Data curation, Programming, Original draft preparation.

Hua Chen: Investigation, Resources, Methodology, Reviewing and Editing, Project administration.

Chong-Yu Xu: Conceptualization Ideas, Methodology, Project administration, Reviewing and Editing.

Ran Huo: Validation, Formal analysis.

Jie Chen: Visualization, Investigation.

Shenglian Guo: Reviewing and Editing.

Declaration of Interest Statement

We declare that we do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted.

Fig. 1. Location and underlying surface mapping of the study area.

Fig. 2. Standardized mean annual precipitation (P), air temperature (T), and mean annual runoff for Daxitan and Xiangxiang basins. The bar chart shows the results of each year, and the red line shows the results of every consecutive 5-year period by moving the window by one year.

Fig. 3. Hierarchical approach for operational testing of hydrological simulations.

Fig. 4. Comparison of the NS and RE values of the five models for the proxy-basin test using the odd and even years. (A = Daxitan; B = Xiangxiang; A_{odd}-B_{even} indicates the calibration in odd years at Daxitan and transfer in even years at Xiangxiang, etc.; Dotted line means $\Delta NS > 0.2$ or $|\Delta RE| > 20\%$)

Fig. 5. Comparison of the NS and RE values of the five models for the differential split-sample test using the driest and wettest consecutive 5-year records. (A = Daxitan; B = Xiangxiang; A_{dry}-A_{wet} indicates the calibration in dry years at Daxitan and transfer in wet years at Daxitan, etc.; Dotted line means $\Delta NS > 0.2$ or $|\Delta RE| > 20\%$)

Fig. 6. Comparison of the NS and RE values of the five models for proxy-basin test using the driest and wettest consecutive 5-year records. (A = Daxitan; B = Xiangxiang; A_{dry}-B_{wet} indicates the calibration in dry years at Daxitan and transfer in wet years at Xiangxiang, etc.; Dotted line means $\Delta NS > 0.2$ or $|\Delta RE| > 20\%$)

Fig. 7. Comparison of the NS value and RE value of the five models using the driest and wettest 1-year records. (A = Daxitan; B = Xiangxiang; A_{dry}-B_{wet} indicates the calibration in dry years at Daxitan and transfer in wet years at Xiangxiang, etc.; Dotted line means $\Delta NS > 0.2$ or $|\Delta RE| > 20\%$)

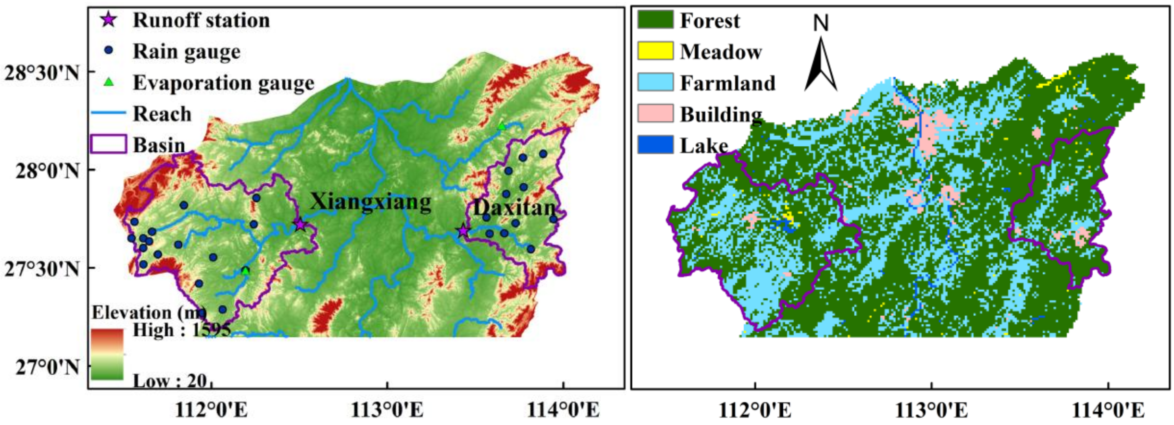

Fig. 1. Location and underlying surface mapping of the study area.

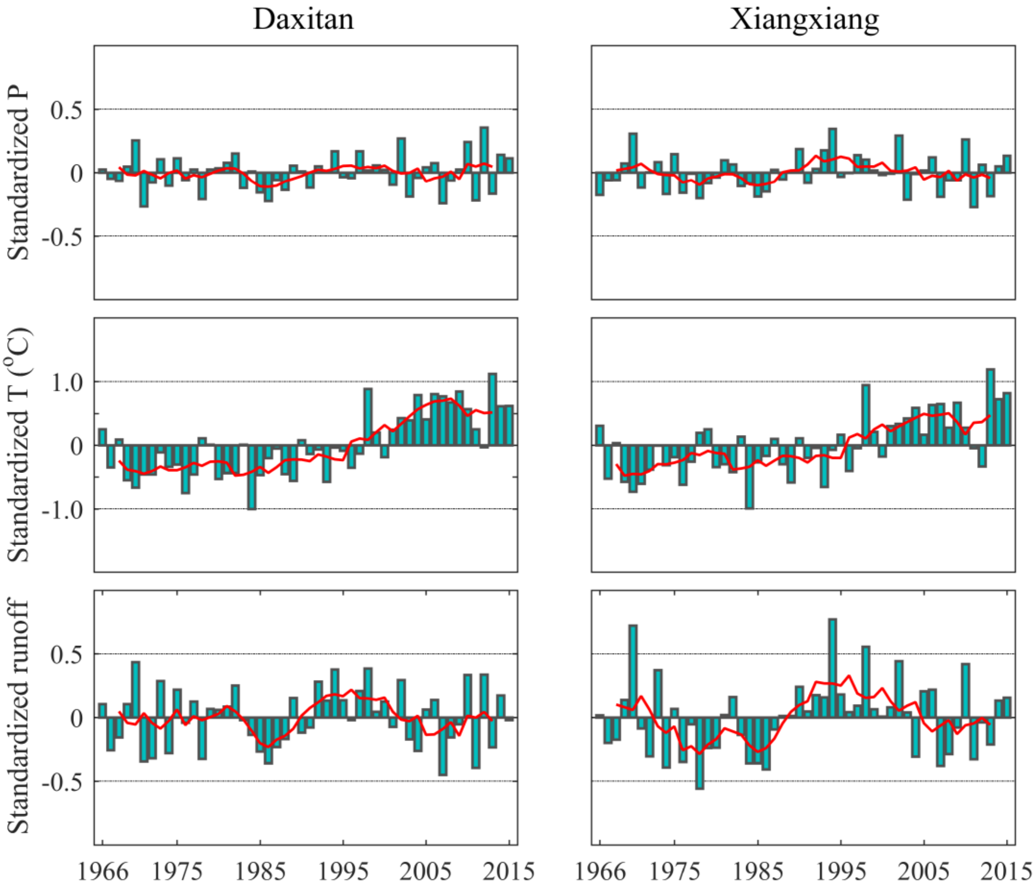

Fig. 2. Standardized mean annual precipitation (P), air temperature (T), and mean annual runoff for Daxitan and Xiangxiang basins. The bar chart shows the results of each year, and the red line shows the results of every consecutive 5-year period by moving the window by one year.

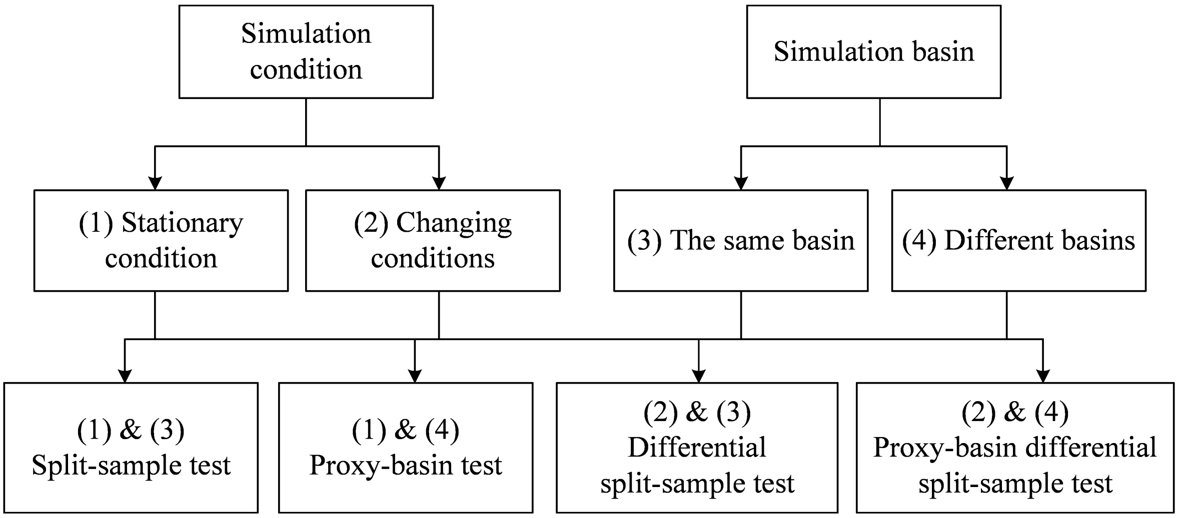

Fig. 3. Hierarchical approach for operational testing of hydrological simulations.

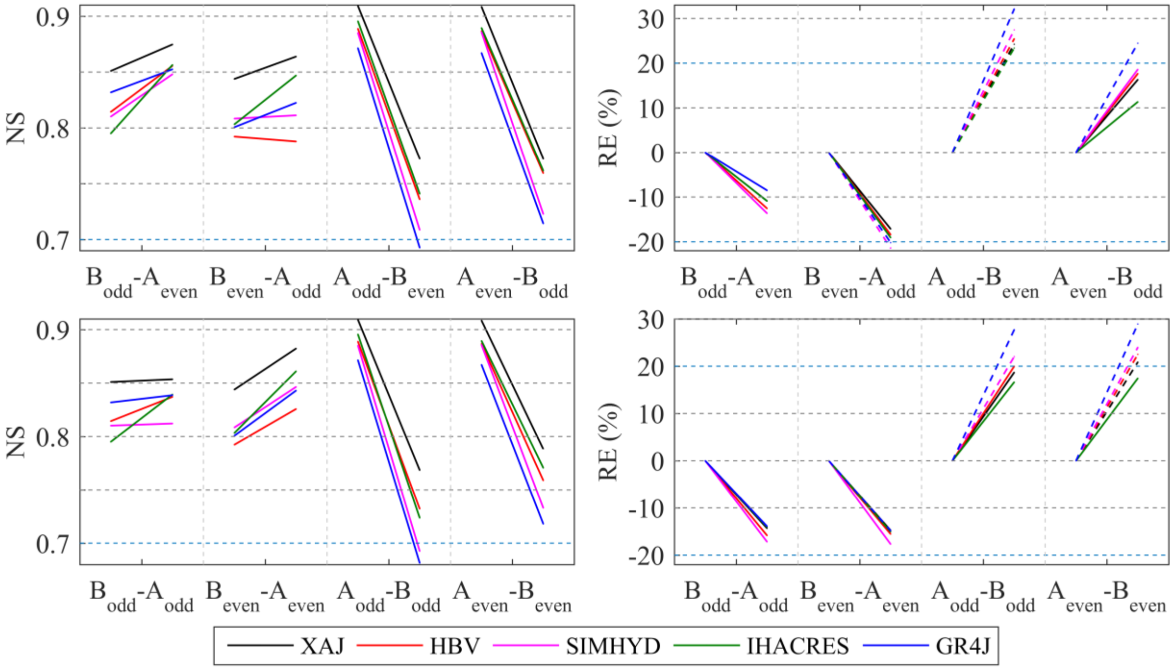

Fig. 4. Comparison of the NS and RE values of the five models for the proxy-basin test using the odd and even years. (A = Daxitan; B = Xiangxiang; A_{odd}-B_{even} indicates the calibration in odd years at Daxitan and transfer in even years at Xiangxiang, etc.; Dotted line means △NS > 0.2 or |△RE| > 20%)

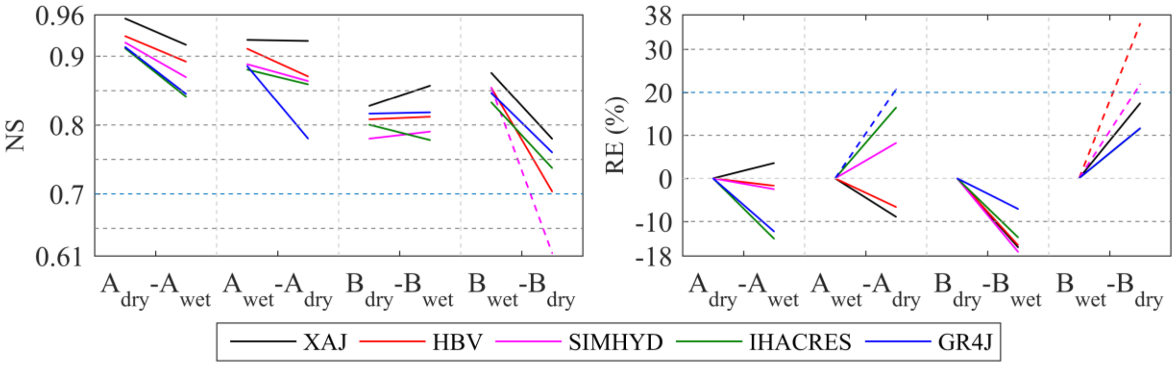

Fig. 5. Comparison of the NS and RE values of the five models for the differential split-sample test using the driest and wettest consecutive 5-year records. (A = Daxitan; B = Xiangxiang; A_{dry}-A_{wet} indicates the calibration in dry years at Daxitan and transfer in wet years at Daxitan, etc.; Dotted line means Δ NS > 0.2 or |Δ RE| > 20%)

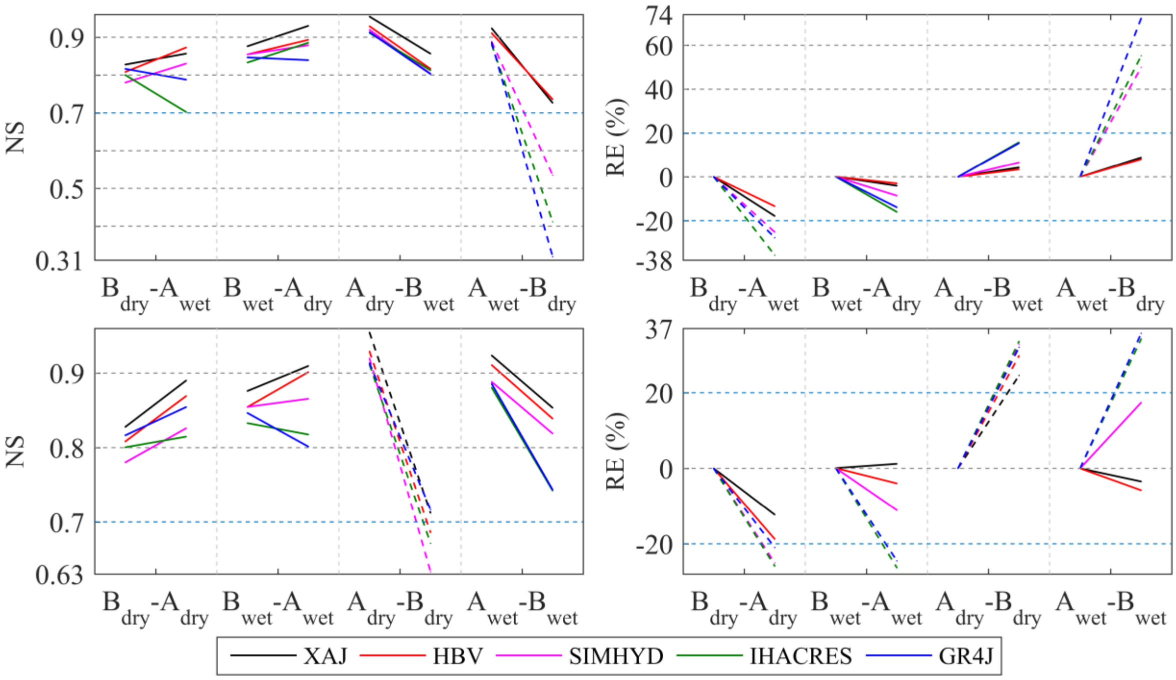

Fig. 6. Comparison of the NS and RE values of the five models for proxy-basin test using the driest and wettest consecutive 5-year records. (A = Daxitan; B = Xiangxiang; A_{dry}-B_{wet} indicates the calibration in dry years at Daxitan and transfer in wet years at Xiangxiang, etc.; Dotted line means $\Delta NS > 0.2$ or $|\Delta RE| > 20\%$)

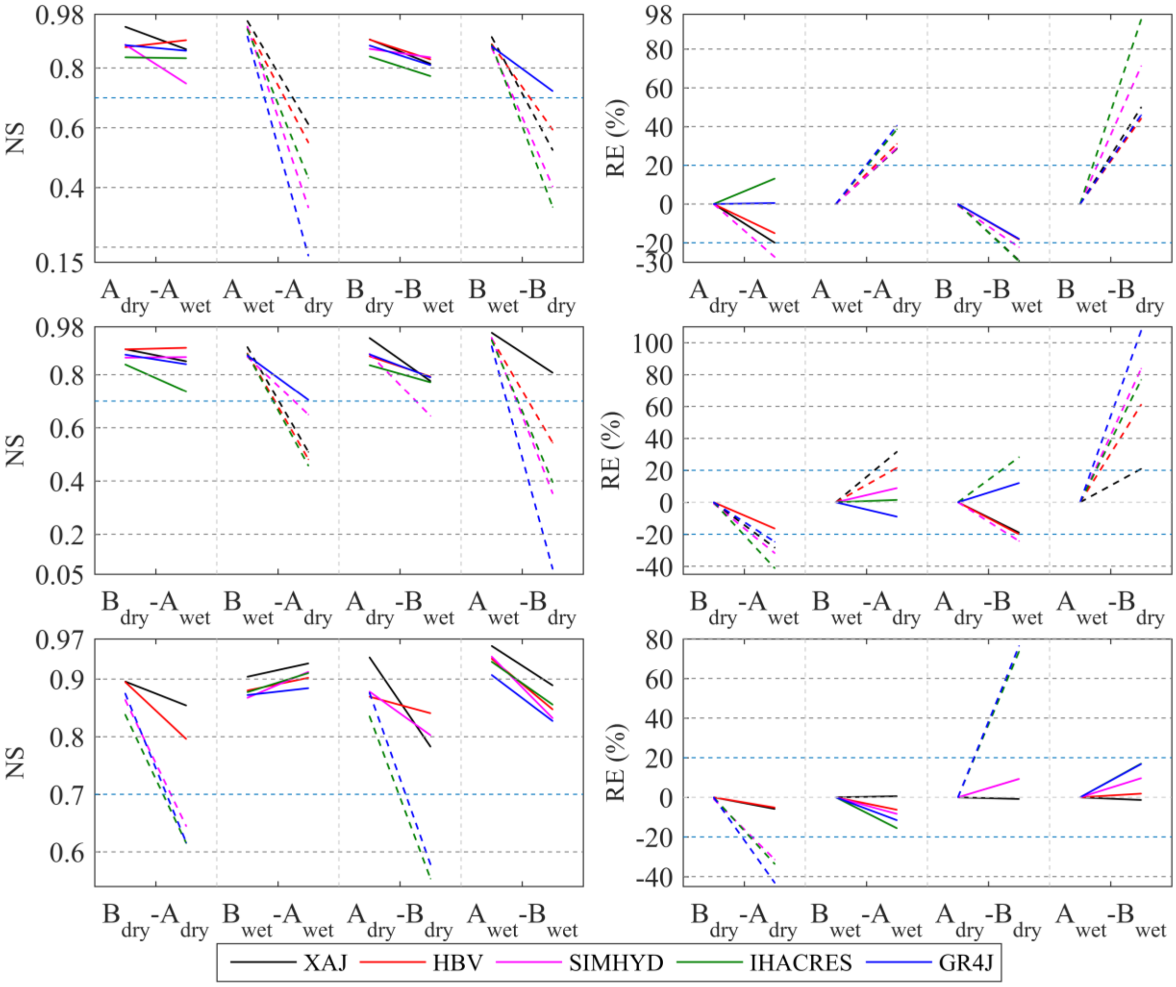

Fig. 7. Comparison of the NS value and RE value of the five models using the driest and wettest 1-year records. (A = Daxitan; B = Xiangxiang; A_{dry}-B_{wet} indicates the calibration in dry years at Daxitan and transfer in wet years at Xiangxiang, etc.; Dotted line means ΔNS > 0.2 or |ΔRE| > 20%)

Table 1. Characteristics of the climate, terrain and vegetation cover of the study basins.

Table 2. Structural components of the five lumped conceptual rainfall-runoff models.

Table 3. Classification of model performance into categories with limits following Moriasi et al. (2007) and Standard of HIHF(Ministry of Water Resources, China, 2008) for NS and RE.

Table 4. Evaluation criteria for hydrological model transferability based on Table 3.

Table 5. Comparison of the statistics results of the five models for temporal transferability test by using odd and even years split-sample test.

Table 6. The hydro-climatic variables with the driest and wettest consecutive 5-year periods.

Table 7. The hydro-climatic variables with the driest and wettest 1-year periods.

Table 8. The sensitivity analysis of the temporal and spatial transferability of the hydrological models under driest year and wettest year.

Table 9. The number of models with transferability under different scenarios.

Table 1. Characteristics of the climate, terrain and vegetation cover of the study basins.

Basin	Area (km^2)	Prec. (mm/m)	Runoff (mm/m)	Runoff Coefficient	Slope (°)	Forest (%)	Lake (%)	Meadow (%)	Farmland (%)	Building (%)
Daxitan	3010	130.1	75.4	0.58	9.60	75.7	0.2	0.4	19.9	3.8
Xiangxiang	5970	113.6	50.2	0.44	8.32	56.5	0.9	0.9	40.5	1.2

Table 2. Structural components of the five lumped conceptual rainfall-runoff models.

Model	Original Authors	Number of Free Parameters	Represented Catchment Stores	Represented Flow Component/ Routing Mechanism
XAJ	Zhao et al. (1980)	15	Upper/Lower/Deep layer tension storage	Surface runoff, Interflow, Groundwater flow; a single uh routing
HBV	Bergström et al. (1976)	10	Soil moisture storage, Groundwater storage	Surface runoff, Interflow, Base flow; a single uh routing
SIMHYD	Chiew et al. (2002)	9	Soil moisture storage, Groundwater storage	Surface runoff, Interflow, Groundwater flow; a single uh routing
IHACRES	Jakeman et al. (1990)	8	Soil moisture storage	Fast flow, Slow flow; a single uh routing
GR4J	Perrin et al. (2003)	4	Production, routing	90% is routed by a uh and then a non-linear routing store, and 10% are routed by a single uh

Table 3. Classification of model performance into categories with limits following Moriasi et al. (2007) and Standard of HIHF(Ministry of Water Resources, China, 2008) for NS and RE.

Sources	Criteria	Performance class			
		Very good	Good	Satisfactory	Unsatisfactory
Moriasi et al. (2007)	NS (-)	> 0.75	0.65-0.75	0.55-0.65	< 0.55
	RE (%)	< 10	10-15	15-20	> 20
Standard of HIHF in China	NS (-)	>0.9	0.90-0.70	0.70-0.50	<0.50
	RE (%)				>20

Table 4. Evaluation criteria for hydrological model transferability based on Table 3.

Criteria			Transferability
NS_T (-) ≥ 0.70	RE_T (%) ≤ 10	-	Acceptable
NS_T (-) < 0.70	RE_T (%) ≤ 10	$\Delta NS \leq 0.2$	Acceptable
		$\Delta NS > 0.2$	Not
NS_T (-) ≥ 0.70	RE_T (%) > 10	ΔRE (%) ≤ 20	Acceptable
		ΔRE (%) > 20	Not
NS_T (-) < 0.70	RE_T (%) > 10	$\Delta NS \leq 0.2$ and ΔRE (%) ≤ 20	Acceptable
		$\Delta NS > 0.2$ or ΔRE (%) > 20	Not

Note: NS_T and RE_T is the NS and RE of the transferred model; ΔNS (ΔRE) is the NS (RE) difference between the calibrated and transferred models.

Table 5. Comparison of the statistics results of the five models for temporal transferability test by using odd and even years split-sample test.

Basin	Period	XAJ		HBV		SIMHYD		IHACRES		GR4J	
		NS	RE (%)	NS	RE (%)	NS	RE (%)	NS	RE (%)	NS	RE(%)
Daxitan	Odd(Cali)	0.91	0.0	0.89	0.0	0.89	0.0	0.90	0.0	0.87	0.0
	Even(Trans)	0.90	3.1	0.88	2.8	0.88	2.7	0.88	5.8	0.87	2.1
	Even(Cali)	0.91	0.0	0.89	0.0	0.89	0.0	0.89	0.0	0.87	0.0
	Odd(Trans)	0.91	-2.1	0.89	3.4	0.88	-3.0	0.89	-1.2	0.87	-2.3
Xiang xiang	Odd(Cali)	0.85	0.0	0.81	0.0	0.81	0.0	0.80	0.0	0.83	0.0
	Even(Trans)	0.84	7.0	0.81	6.4	0.80	5.5	0.79	9.6	0.80	8.7
	Even(Cali)	0.84	0.0	0.79	0.0	0.81	0.0	0.80	0.0	0.80	0.0
	Odd(Trans)	0.84	-2.6	0.80	9.5	0.81	-5.6	0.80	-4.6	0.83	-8.0

Table 6. The hydro-climatic variables with the driest and wettest consecutive 5-year periods.

Basin	Variables	Dry	Wet	Variability
Daxitan	P (mm/month)	116.0	137.4	18.4%
	T (°C)	17.2	17.6	0.4°C
	Q (mm/month)	60.1	91.8	52.7%
	Runoff coefficient	0.52	0.67	29.0%
Xiangxiang	P (mm/month)	102.7	121.6	18.4%
	T (°C)	16.9	16.6	0.3°C
	Q (mm/month)	35.8	58.7	64.0%
	Runoff coefficient	0.35	0.48	38.5%

Table 7. The hydro-climatic variables with the driest and wettest 1-year periods.

Basin	Variables	Dry	Wet	Variability
Daxitan	P (mm/month)	98.8	163.2	65.2%
	T (°C)	18.4	16.9	1.5°C
	Q (mm/month)	41.4	108.3	161.6%
	Runoff coefficient	0.42	0.66	58.4%
Xiangxiang	P (mm/month)	90.7	148.7	64.4%
	T (°C)	17.3	16.3	1.0°C
	Q (mm/month)	22.1	86.4	291.0%
	Runoff coefficient	0.24	0.58	138.5%

Table 8. The sensitivity analysis of the temporal and spatial transferability of the hydrological models under driest year and wettest year.

Transferability	No. of Scenarios	XAJ	HBV	SIMHYD	IHACRES	GR4J	Test Count	Accepted count
A_{dry}-$A_{5\%25\%50\%75\%95\%}$	5	5	5	4	5	5	25	24
A_{wet}-$A_{5\%25\%50\%75\%95\%}$	5	5	5	5	4	3	25	22
B_{dry}-$B_{5\%25\%50\%75\%95\%}$	5	3	2	2	1	4	25	12
B_{wet}-$B_{5\%25\%50\%75\%95\%}$	5	3	1	1	0	3	25	8
A_{dry}-$B_{5\%25\%50\%75\%95\%}$	5	1	3	4	2	1	25	11
A_{wet}-$B_{5\%25\%50\%75\%95\%}$	5	3	2	2	2	2	25	11
B_{dry}-$A_{5\%25\%50\%75\%95\%}$	5	5	4	0	0	0	25	9
B_{wet}-$A_{5\%25\%50\%75\%95\%}$	5	5	4	4	0	4	25	17
Total	40	30	26	22	14	22	200	114

Note: A-Daxitan; B-Xiangxiang

Table 9. The number of models with transferability under different scenarios.

Transferability	Scenarios Number	XAJ	HBV	SIMHYD	IHACRES	GR4J	Test Count	Accepted count	Percentage (%)
Total	76	58	54	40	34	40	380	224	58.9
Temporal	32	26	22	19	19	24	160	110	68.8
Spatial	12	10	10	7	6	4	60	37	61.7
Temporal-Spatial	32	22	22	14	9	12	160	79	52.7
A→B	22	12	13	11	9	6	110	51	46.4
B→A	22	20	19	10	6	10	110	65	59.1
Dry→Wet	8	7	8	3	4	6	40	28	70.0
Wet→Dry	8	4	3	2	3	3	40	15	37.5

Note: A-Daxitan, B-Xiangxiang; **Temporal:** A(B)$_{\text{odd/even/dry5/wet5/dry1/wet1}}$-A(B)$_{\text{even/odd/wet5/dry5 /wet1/dry1}}$，A(B)$_{\text{dry1/wet1}}$-A(B)$_{\text{5\%,25\%,50\%,75\%,95\%}}$； **Spatial:** A(B)$_{\text{odd/even/dry5/wet5/dry1/wet1}}$-B(A)$_{\text{odd/even/dry5/wet5/dry1/wet1}}$； **Temporal-Spatial:** A(B)$_{\text{odd/even/dry5/wet5/dry1/wet1}}$-B(A)$_{\text{even/odd/wet5/dry5/wet1/dry1}}$；A(B)$_{\text{dry1/wet1}}$-B(A)$_{\text{5\%,25\%,50\%,75\%,95\%}}$; **A(B)→B(A):** A(B)$_{\text{odd/even/dry5/wet5/dry1/wet1}}$-B(A)$_{\text{odd/even/dry5/wet5/dry1/wet1}}$，A(B)$_{\text{odd/even/dry5/wet5/dry1/wet1}}$-B(A)$_{\text{even/odd/wet5/dry5/wet1/dry1}}$；A(B)$_{\text{dry1/wet1}}$-B(A)$_{\text{5\%,25\%,50\%,75\%,95\%}}$； **Dry→Wet:** A(B)$_{\text{dry5/dry1}}$-A(B)$_{\text{wet5//wet1}}$，B(A)$_{\text{dry5/dry1}}$-A(B)$_{\text{wet5//wet1}}$； **Wet→Dry:** A(B)$_{\text{wet5/wet1}}$-A(B)$_{\text{dry5/dry1}}$，B(A)$_{\text{wet5/wet1}}$-A(B)$_{\text{dry5/dry1}}$