Comparing spatial and temporal scales of hydrologic model parameter transfer: A guide to four climates of Iran

Jahanshahi, Afshin; Melsen, Lieke; Patil, Sopan; Goharian, Erfan

Journal of Hydrology

DOI:
https://doi.org/10.1016/j.jhydrol.2021.127099

Published: 01/12/2021

Peer reviewed version

Comparing spatial and temporal scales of hydrologic model parameter transfer: A guide to four climates of Iran

Afshin Jahanshahi*[1] Lieke A. Melsen[2] Sopan D. Patil[3] Erfan Goharian[4]

1- Department of Watershed Management, Sari Agricultural Sciences and Natural Resources University (SANRU), P.O. Box 737, Sari, Iran.

Corresponding author (Email: a.jahanshahi@stu.sanru.ac.ir). Tel: +98 11 3388 2981.

2- Hydrology and Quantitative Water Management Group, Wageningen University, Droevendaalsesteeg 3a, 6708 PB Wageningen, the Netherlands.

3- School of Natural Sciences, Bangor University, Deiniol Road, Bangor LL57 2UW, United Kingdom.

4- Department of Civil and Environmental Engineering, University of South Carolina, Columbia, SC 29208, USA.

Abstract

Simulating streamflow in ungauged catchments remains a challenging task in hydrology and increases the demand for regionalization studies worldwide. Here, we investigate the effect of three modes of parameter transfer, including temporal (transferring across different periods), spatial (transferring between same calibration periods but different sites), and spatiotemporal (transferring across both different periods and sites) on simulating streamflow using HBV conceptual rainfall-runoff model at 576 unregulated catchments throughout Iran (407,000 Km^2). Our main conclusions are: (1) temporal mode shows the best performance, with the lowest decline in performance (median decline of 5.8%) as measured using the NSE efficiency metric, (2) difference between spatial and spatiotemporal options was negligible (median decline of 13.7% and 15.1% respectively), (3) all parameters are associated with some uncertainties and those related to runoff and snow components of the model are associated with the highest and lowest uncertainties, respectively, (4) overall, the model performance in arid regions is not as good as humid regions which confirmed that elevation and climate play a major role in parameter estimation in these areas, and (5) aridity and catchment elevation are two major controls on model transferability at regional (climate classes) and local (the whole country) scales. We also show that the superiority of the temporal mode is maintained with: (i) increasing spatial distance between gauged (donor) and ungauged (target) catchments, (ii) increasing time lag (10 years) between calibration and validation, and (iii) gradually increased time lags between calibration and validation. Our study suggest that spatiotemporal parameter transfer can be a reliable option for PUB studies and climate change-related studies, at least in wetter catchments. However, further research is needed to explore the complicated relationship between temporal and spatial aspects of hydrological variability.

Keywords: Aridity, Parameter transfer, Rainfall-runoff model, Ungauged catchment

Introduction

The simulation of streamflow in ungauged catchments remains a challenging task in the hydrologic sciences (Sivapalan et al., 2003) because the model parameters cannot be calibrated against streamflow since there are no observations. The process of finding appropriate parameter sets to simulate streamflow in ungauged sites by learning from model calibration in gauged sites is generally referred to as "regionalization". Over the last decade, an increasing number of studies have used conceptual rainfall-runoff models to test different regionalization approaches (Lee et al., 2005; Merz and Blöschl, 2004; Perrin et al., 2001; Reichl et al., 2009; Samuel et al., 2011; Vaze et al., 2010; Vogel, 2005; X. Yang et al., 2020a; Yang et al., 2019). The main issue in regionalization is related to the operational application of these models outside of calibration periods, where the parameter sets face their true examination (Dakhlaoui et al., 2017; Patil and Stieglitz, 2015; Refsgaard and Knudsen, 1996; Yang et al., 2020b). Parameter transfer, or regionalization, outside of the calibration period can be in time (simulating streamflow for periods for which no observations are available), in space (simulating in ungauged sites) or both (hereafter referred to as "spatiotemporal"). Hrachowitz et al. (2013), Blöschl et al. (2013), and Parajka et al. (2013) provide a comprehensive overview of the achievements and discussions in PUB research during the PUB decade initiative (2003-2012) initiative of the International Association of Hydrological Sciences (IAHS).

Temporal transfer of hydrological model parameters is the most common approach in regionalization studies (Patil and Stieglitz, 2015). An implicit assumption in the temporal transfer is that calibrated parameters are temporally stable. However, many recent studies have shown that calibrated model parameters have not been temporally stable (e.g., Brigode et al., 2013; Dakhlaoui et al., 2017; Merz et al., 2011; Yang et al., 2018) and conditions of calibration period determine their values (Juston et al., 2009).

Parameters transfer in spatial mode, from gauged to ungauged sites, is another strategy widely used in numerous studies for streamflow prediction in ungauged basins (PUB) across the world (Choubin et al., 2019; McIntyre et al., 2005; Oudin et al., 2008; Patil and Stieglitz, 2014, 2015; Samuel et al., 2011; Sivapalan et al., 2003; Yang et al., 2018; Young, 2006; Zhang and Chiew, 2009). Two widely used approaches are Spatial Proximity (SP) and Physical Similarity (PS). The implicit hypothesis of the SP approach is that two adjacent catchments behave similarly in hydrological response because they are likely to have similar physical and climatic conditions (Chiew et al., 2008; Petheram et al., 2009). However, this may not always be the case as nearby catchments can sometimes have different characteristics and therefore not behave similarly (Kennard et al., 2010; Petheream and Bristow, 2008; Thornton et al., 2007). The second parameter transfer approach is PS (Choubin et al., 2019; Kay et al., 2007; Samaniego et al., 2010). In the PS transfer approach, parameter sets are transferred from the most physically similar catchment(s) to the ungauged catchment (Bao et al., 2012; Bárdossy, 2007; McIntyre et al., 2005; Samuel et al., 2011). It remains, however, challenging to determine which physical characteristics are key for successful parameter regionalization.

A few studies have pointed to a significant difference between the performance of temporal and spatial parameter transfer (e.g., Arsenault and Brissette, 2014; Merz and Blöschl, 2004; Parajka et al., 2005; Yang et al., 2020a; Zhang and Chiew, 2009) and some have shown less difference between their performance (Oudin et al., 2008; Patil and Stieglitz, 2015). Although there is a considerable number of PUB studies on the development and comparison of approaches to transfer rainfall-runoff model parameters from gauged to ungauged catchments in different sites, both in terms of size and climate (McIntyre et al., 2005; Merz et al., 2011; Post and Jakeman, 1996; Young, 2006), not many studies have carried out a direct comparison of these three modes within/between climate classes in a study area. The present paper goes beyond Patil and Stieglitz (2015) in terms of climate classes and the number of studied catchments and also adds to the existing literature by addressing the PUB paradox. Many PUB studies are conducted in catchments where actually many observations are available. Consequently, these studies, especially those with a high number of catchments involved, have been conducted in regions with a dense and well-organized observation network (the US, Austria, and France), mainly temperature climates, where the need for regionalization might be limited. Here, we investigate several methods for regionalization across Iran. Iran is an example of a country where regionalization might be even more important because of the low gauge density and a climate regime that is completely different compared to the widely studied in France, Austria, and the US.

In Iran, streamflow gauges are generally not in good condition. There are only 1,194 active gauges and with respect to the total area of the country (1,648,000 km^2), there is only one active gauge per 1,380 Km^2 (IEM, 2016). The minimum density of streamflow gauges recommended by WMO (WMO, 2009) is one gauge per 1,875 km^2 and 1,000 km^2 for mountains and interior plains, respectively. Prediction of streamflow time series in ungauged catchments is a global challenge in hydrology, and this also applies to Iran - especially in its arid and semi-arid regions. Hence, PUB is chosen as an essential issue in this study, where we utilize three modes of temporal, spatial, and spatiotemporal parameter transfer using the available dataset in Iran that covers an extensive range of climate types.

To our knowledge, there are only two PUB studies conducted in Iran across the Karkheh River Basin in the west. Masih et al. (2010) defined hydrological similarity based on four similarity measures: spatial proximity, drainage area, catchment properties, and Flow Duration Curves (FDC) in 11 ungauged catchments. Their results showed that the physical similarity approach based on similarity in quantiles of FDC in the HBV model leads to the best performance. In another study, Choubin et al. (2019) defined the catchment similarity based on morphological, topographic, soil type and land use, and remote sensing-based characteristics in four catchments. They concluded that physical similarity by applying the semi-distributed SWAT model is an efficient method to estimate streamflow times series in ungauged catchments.

Our study compares temporal transferability with spatial and spatiotemporal strategies, using the HBV hydrological model across 576 catchments throughout Iran (Fig. 1). The temporal mode is implemented using a split-sample test procedure (Parajka et al., 2005), where the available data is divided into two calibration and validation periods. We use the nearest neighbor catchment as a donor of calibrated parameters for the spatial/spatiotemporal parameter transfer strategy.

The main questions addressed in this study are:

(i) How do spatial and spatiotemporal transfer of hydrological model parameters differ with an increase in lag time (temporal) between calibration and validation?

(ii) How do spatial and spatiotemporal transfer of the hydrological model differ with an increase in spatial distance between gauged (donor) and ungauged (target) catchments?

(iii) How do hydrological model parameters differ between two calibration periods?

(iv) How do dynamic and statistic catchment characteristics control model transferability at local and regional scales?

2. Study area, model, and dataset

2.1. Study area

Our study area is Iran. There are four general climate regions in Iran based on De Martonne classification system (De Martonne, 1926; Rahimi et al. 2013). The climate varies greatly within the country, from wet maritime weather along the Caspian Sea coast, including humid and semi-humid, toward drier conditions in the interior, including arid and semi-arid.

There is considerable annual and seasonal variability in rainfall across Iran, with mean annual precipitation (MAP) ranges from 360 mm (central parts) to more than 2000 mm (northern parts of the country) (mean = 724 mm) (IMO, 2018). The spatial variability of precipitation is particularly large between the north, northwest, west, and central parts of the country (from less than 400 to more than 2000 mm) (see Table 2). Altitude greatly impacts the amount of rainfall in the mountainous areas of Iran, and runoff hydrographs show quite different spatial patterns (IEM, 2018).

2.2. HBV rainfall-runoff model

The HBV model is a semi-distributed conceptual rainfall-runoff model. It was originally developed in Sweden (Bergström, 1976). It requires three input variables at the daily time step: precipitation, temperature, and potential evapotranspiration. For PUB studies, it has been widely used in semi-arid (Choubin et al., 2019; Lidén and Harlin, 2000; Love et al., 2010; Masih et al., 2010)and humid (Clark et al., 2017; Merz and Blöschl, 2004; Pool et al., 2017; Samuel et al., 2011; Seibert and Beven, 2009) regions. The model version used herein, modified by Parajka and Viglione (2012), includes snow routine, a soil routine, routing routine using the unit hydrograph, and a response function with three linear reservoir equations (Osuch et al., 2019). This modified version has 15 parameters (Table 4) (Parajka et al., 2007).

2.3. Forcing data

Daily precipitation time series for all catchments are aggregated from the Iran precipitation dataset provided by the Iran Energy Ministry (IEM) (IEM, 2018) and Iran Meteorological Organization (IMO) (IMO, 2018). In this dataset, rainfall data are collected from point observations at gauge locations, but we estimated rainfall fields through two methods, IDEW and lapse rate:

(i) IDW and Elevation (IDEW) method. The IDEW is an interpolation technique and offers the possibility of defining elevation and distance weighting, making it more suitable for mountainous regions of Iran. This technique was shown to be more suitable for mountainous catchments in the

Karkheh River Basin and southwestern Iran (Masih et al., 2011, 2010; Modallakdoust et al., 2008). The equation for this method is as follows:

$$\hat{p}_k = W_D \sum_{i=1}^{N} \frac{1}{D} w(d)_i \, p_i + W_Z \sum_{i=1}^{N} \frac{1}{Z} w(z)_i \, p_i \qquad (1)$$

where $\hat{p}_k$ is interpolated precipitation for grid cell (mm/time step), W_z(-) and W_D (-) are total weighting factors for elevation and distance, respectively, p_i is precipitation value (mm/time step) of the i-th gauge station, and N is the number of precipitation gauges used for interpolation of the current grid cell. Similarly, $w(z)_i$ (–) and $w(d)_i$ (–) are the individual gauge weighting factors for elevation and distance, respectively, and Z (–) and D (–) are the normalization quantities given by the sum of individual weighting factors $w(z)_i$ and $w(d)_i$, respectively, for all interpolated gauges. The weighting factors $w(d)_i$ and w(z) based on the elevation and inverse of distance are as follows:

$$w(d) = 1/d^a \quad \text{for} \quad d > 0 \qquad (2)$$

$$w(z) = \begin{cases} 1/\mathrm{Z}_{\min}^{b} & \text{for } z \le z_{\min} \\ 1/\mathrm{z}^{b} & \text{for } z_{\min} < z < z_{\max} \\ 0 & \text{for } z \le z_{\min} \end{cases} \qquad (3)$$

where d is the distance (km) between the current grid and the precipitation gauge, z is the absolute elevation difference (m) between the current grid cell and the precipitation gauge, b (–) and a (–) are constants for elevation and distance weightings, respectively, and z_{max} (m) and z_{min} (m) are the maximum and minimum limiting values for computing elevation weightings.

Time series of daily precipitation data are used for interpolation in 5 × 5 km^2 grids, which are then aggregated at the catchment scale. The parameters of interpolation, i.e., the exponents a and b, the radius of influence, and importance factors W_Z and W_D and, are determined by cross-validating the interpolated precipitation using Jack-Knife method (Varljen et al., 1999). The cross-validation was done for 1081 selected grid cells/precipitation gauge locations scattered throughout Iran.

The monthly R^2 (coefficient of determination) ranges from 0.58 to 0.92. Considering the high spatial variability of precipitation in highlands, the R^2 values are considered satisfactory. A detailed comparison of model efficiency under areal precipitation and gauge (point observations) data is beyond the scope of this paper. The main parameters used for the interpolation were: $W_D = 0.8$ and $W_Z = 0.2$, radius of influence = 80 km, a = 2, and b = 1 (Masih et al., 2010). The limiting values for elevation weighting z_{max} and z_{min} are selected as 4600 m and 40 m, respectively.

(ii) A lapse rate correction ~~is~~ was also tested to account for the elevation effect. Catchment rainfall is increased by a correction factor that is allowed to vary with mean catchment elevation. This correction factor was set to 4% in lowland catchments of Iran (below 150 m a.s.l.) and 18 to 25% in the highland catchments (above 2000 m a.s.l.) and assumed to increase linearly in between.

The results showed that the difference between the IDEW and the lapse rate methods, in terms of catchment averaged rainfall is relatively small (in the order of 3.8%). Hence, the values of catchment rainfall calculated by the IDEW method are used for this study.

There is at least one rain gauge and temperature/evaporation station in each catchment. Daily temperature time series are generated from the IEM and IMO observations using a regression-based method by applying elevation as explanatory factor. The reference evapotranspiration is estimated with the Hargreaves method (Hargreaves et al., 1985) using maximum, minimum, and average temperature. With the Standard Normal Homogeneity Test (SNHT) (Haimberger, 2007), the time series of three model inputs was shown to be homogenous, and no breakpoints were observed. The missing values in the data sets were estimated based on the values from neighboring gauges using the regression method. Overall, the temperature data of a gauge showed a good correlation with corresponding data from the neighboring gauges ($R^2 > 0.89$) used for filling the missing records. In the case of precipitation data, this correlation is $R^2 > 0.85$. On average, 7.2% and 10.5% of the temperature and precipitation data, respectively, had to be filled for all 963 catchments.

Figure 1 shows the climatic attributes estimated using the period from 1993 to 2008 for the 576 study catchments. This figure shows low climate variability in precipitation, temperature, PET, and percentage of snow cover between two calibration periods in the study catchments. There is no trend or change point in the annual mean of four climatic variables from 1993 to 2008 in the study catchments, evaluated with the Hubert's segmentation procedure. Thus, variability in four variables is in the form of inter-annual. Table 1 shows the mean annual values of climatic variables for four climate regions over two calibration periods (1993-2001 and 2001-2008).

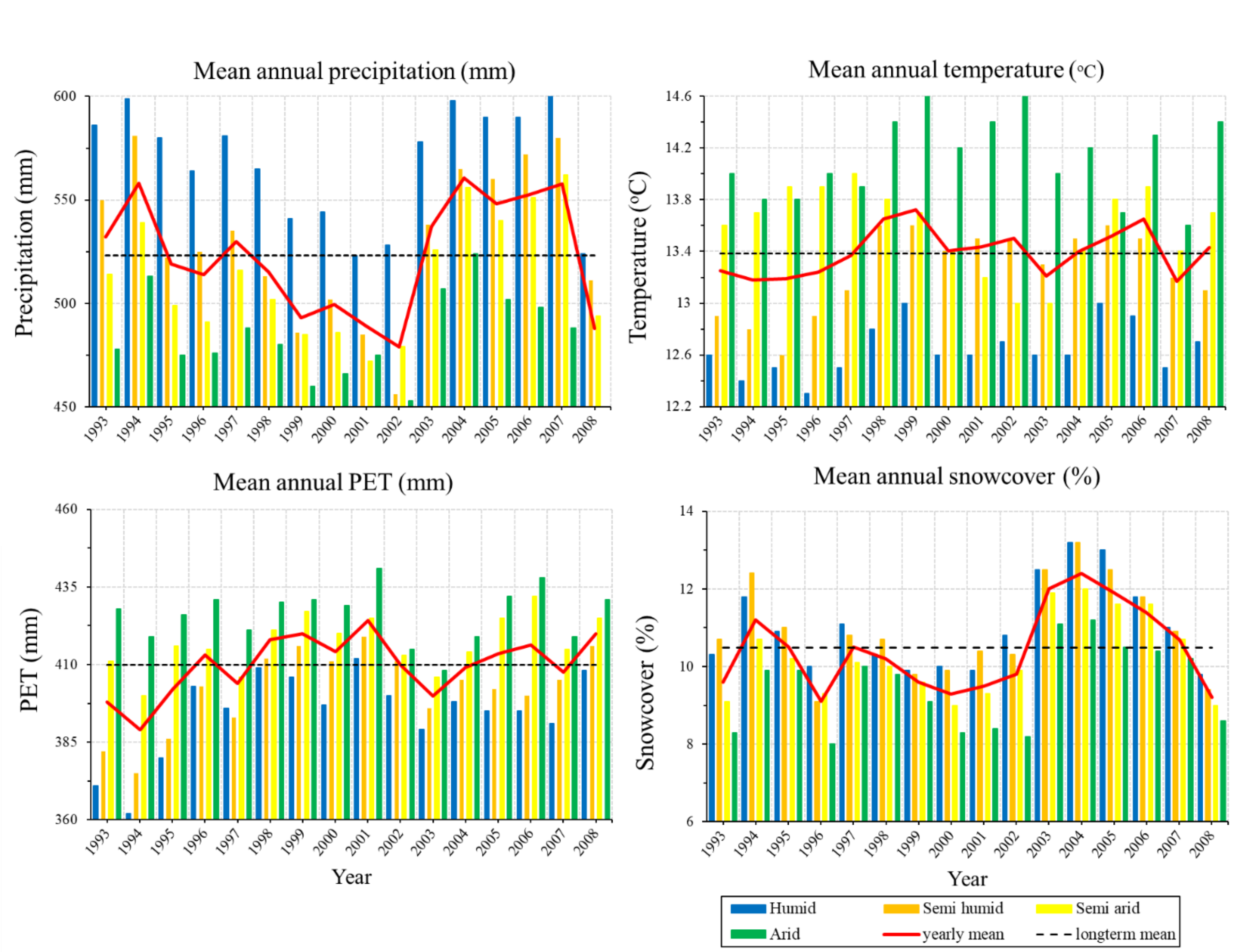

Fig. 1. Inter-annual climate variability over the period 1993 to 2008 in the study catchments. The period is split into two: 1993-2001 and 2001-2008.

Table 1

Mean annual values of four catchment characteristics for four climate regions over two calibration periods.

Region	Calibration Period	Precipitation (mm)	Temperature (ºC)	PET (mm)	Snow cover (%)
Humid	1993-2001	570	12.58	390.5	10.5
	2001-2008	566	12.7	390.8	11.5
Semi humid	1993-2001	526	13.11	397.25	10.6
	2001-2008	533.3	13.4	406.75	11.3
Semi-arid	1993-2001	504	13.75	414.5	9.8
	2001-2008	522.5	13.42	419.37	10.7
Arid	1993-2001	479.5	14	427	9.1
	2001-2008	483.7	14.17	425.3	9.8
Iran (all regions)	1993-2001	520	13.37	407.27	10
	2001-2008	526	13.41	412.5	10.9

2.4. Catchment dataset

The dataset used in this study includes daily precipitation at 1081 stations and daily air temperature at 612 climatic stations in 996 catchments. Digital maps of land use (MODIS Land Cover Product), global soil map (based on the FAO map), aquifers map (based on 1:250000 Iran energy Ministry map), and the main geological formations (1:250000 map of USGS) are used. These digital maps are combined with the catchment boundaries to derive each land-use type, soil type, aquifer area, and geological unit. 33 catchments out of 996 preliminary catchments with high permeability (karstic aquifers) and dry catchments with very high and variable permeability are removed since the employed model is not capable to simulate these conditions. We carefully screened the runoff data for errors, and outliers are removed.

To calibrate and validate the HBV model, daily runoff data from 963 gauged catchments are used with areas ranging from 64.7 km^2 to 8432 km^2 and a median of 496 km^2 (Table 2). 97 of these catchments range in area between 64.7 and 150 km^2, 206 catchments between 150 and 350 km^2, 334 catchments between 350 and 1000 km^2, 326 catchments between 1000 and 8432 km^2. The catchment area increases from wet to dry regions so that its median values for humid and semi-arid regions are the lowest and highest, respectively (Table 3). The spatial distribution of measuring gauges (in terms of number and distribution per unit area) deteriorates from wet to dry conditions for hydrometric and meteorological stations. This catchment dataset has a daily streamflow dataset from water year (WY) 1992 to 2008 (i.e., September 22, 1992, to September 21, 2008). In this study, the period of 1993 to 2008 is split into two consecutive calibration periods. Catchment descriptors for all 963 study catchments are presented in Table 2.

The water years 1993-2001 were used as calibration period 1 and the water years 2001-2008 for calibration period 2. The period 1992-1993 is used for model warm-up. We calibrated HBV model parameters for these two periods separately. Only those catchments where the results provide NSE ≥ 0.5 (following Parajka et al., 2007) for both periods are retained for the next step. Applying this threshold reduced the number of catchments considered in this study to 576 (Fig. 2). The minimum and median catchment areas are 104.6 km^2 and 608.2 km^2, respectively. Figure 2 shows the location of 576 selected catchments and their classification into four climate regions. Table 3 shows the median values of catchment descriptors for different climate regions.

Table 2

Statistics of catchment descriptors ~~(CDs)~~ (n = 963).

Category	Catchment descriptor	Median	Range	
			Min	Max
Topographic	Mean elevation (m)	1416	-28	5,595
	Mean slope (%)	18.8	0	45.1
Physiographic	Area (km^2)	574.2	64.7	8,432
Climatic	Aridity Index (PET/P) (-)	0.69	0.1	2.83
	Mean annual precipitation (mm)	724	367	2015
	Mean annual temperature (°C)	14.3	5.9	24.4
	PET (mm)	459	151	1194
Land use	Rangeland (%)	33.1	2.8	87.4
	Agriculture (%)	27.6	3.42	46.42
	Forest (%)	13.5	0.01	41.7
	Residential (%)	4.3	0	21.8

Table 3

The median values of catchment descriptors for different climate regions (n = 576).

Catchment descriptor	Humid (G1)	Semi-humid (G2)	Semi-arid (G3)	Arid (G4)
No. of catchments	199	256	93	28
Area (km^2)	372	538	1192	1096
Mean elevation (m)	1776	2426	873	368
Mean slope (%)	22.5	31.2	15.3	10.2
Aridity Index (-)	0.35	0.49	0.59	1.24
Mean annual precipitation (mm)	1065	816	673	392
Mean annual temperature (°C)	8.4	10.2	14	20
PET (mm)	293	390	386	718
Rangeland (%)	15.4	26.1	37.2	52.6
Agriculture (%)	19.3	28.7	32.7	27.6
Forest (%)	23.8	16.9	9.2	3.3
Residential (%)	3.7	6.1	5.2	2.4

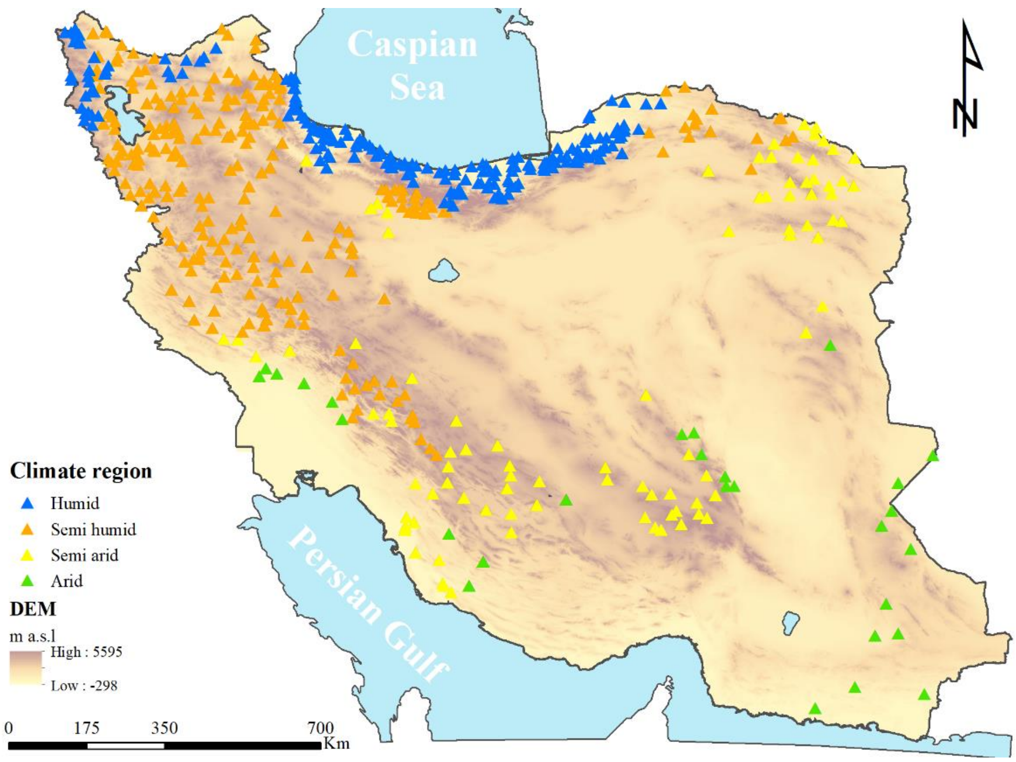

Fig. 2. The location of 576 catchments grouped in climate regions; blue is humid, orange is semi-humid, yellow is semi-arid, and green is arid. The triangles have been plotted at the outlet of catchments.

3. Methodology

3.1. Model calibration and evaluation

Adjusting hydrological model parameters is an essential part of hydrological simulations. The goodness-of-fit was improved by optimizing these parameter values until the difference between measured and simulated runoff was satisfactory during model calibration.

The Differential Evolution optimization algorithm (DEoptim) (Storn and Price, 1997) is used to calibrate the model parameters through the DEoptim package in R (Ardia et al., 2011). The algorithm is in the class of genetic algorithms that maximize a given objective function (Mitchell, 1998). DEoptim parameters were set to itermax = 400, population size (NP) = 400, trace = 7, crossover probability = 0.5, and step-size = 0.8. The upper and lower boundaries of each HBV parameter were determined according to Parajka et al. (2007) (Table 4). The model was run in a lumped fashion for each catchment.

Model calibration and evaluation (transferability) are evaluated using Nash-Sutcliffe Efficiency (NSE) (Nash and Sutcliffe, 1970). The NSE criterion is a form of the normalized least-squares objective function. It places more emphasis on high flows. Its optimal value is 1.

$$NSE=\left[\frac{\sum_{i=1}^{n}(Q_i^{obs}-Q_i^{sim})^2}{\sum_{i=1}^{n}(Q_i^{obs}-\overline{Q_{obs}})^2}\right] \quad (4)$$

where Q_i^{sim} and Q_i^{obs} are the daily simulated and observed runoff values at the time i, respectively. $\overline{Q_{obs}}$ is the mean value of daily observed runoff. One year before each calibration period was used as the warm-up period to reduce the impact of the uncertain initial conditions on model performance.

3.2. Parameter transfer strategies and tested modes

In this study, parameter transfer from gauged to ungauged catchments is examined under spatial and spatiotemporal strategies. Model performance in gauged catchments is also examined under temporal mode. Patil and Stieglitz (2015) also examined these tested strategies and mode in 294 catchments in the United States. The spatial strategy used in this study is the Nearest Neighbor (NN) method. This spatial parameter transfer strategy is carried out at the scale of regional (climate regions) and local (entire Iran). We examine them across Iran as follows:

(1) *Temporal* (TEM): Model parameters from period 1 are transferred to period 2, and vice versa (for the same catchment).

(2) *Spatial* (SPA): Model parameters of a catchment are obtained from transferring method (nearest neighbor) over the same time period (separately for two calibration periods).

(3) *Spatiotemporal* (SPA_TEM): Here, model parameters are transferred in temporal (between time periods) and spatial (by nearest neighbor) domain, simultaneously.

4. Results

4.1. Model performance over the calibration periods (at-site)

We calibrate the model for 963 catchments in two consecutive calibration periods. None of them showed an NSE lower than 0.29. Subsequently, we eliminate 349 from the initial set of 963 unregulated catchments due to poor model performance for two consecutive periods (calibration NSE < 0.5). The discarded catchments (387 catchments with calibration NSE < 0.5) are mostly in dry (n = 218) and some in wet (n = 169) regions, where karstic aquifers show complex interaction between surface water and groundwater. Overall, 576 gauged catchments out of the 963 catchments set showed NSE ≥ 0.5 for both periods (Fig. 3 left panel). These results, in combination with Fig. 1, show that the model provides better performance in wetter catchments (humid and semi-humid) than the drier ones (arid and semi-arid) and are generally in accordance with the findings by Parajka et al. (2005) in Austria and Oudin et al. (2008) in France.

Figure 3 (right panel) shows the spatial distribution of average model performance for the 576 catchments over calibration periods. In the northwestern, northern, and across the interior western catchments, model performance is substantially better than in the other parts of the country. Conversely, interior, western and southeastern catchments are generally more difficult to model since spatially variable rainfall events make the streamflow vary in amplitude.

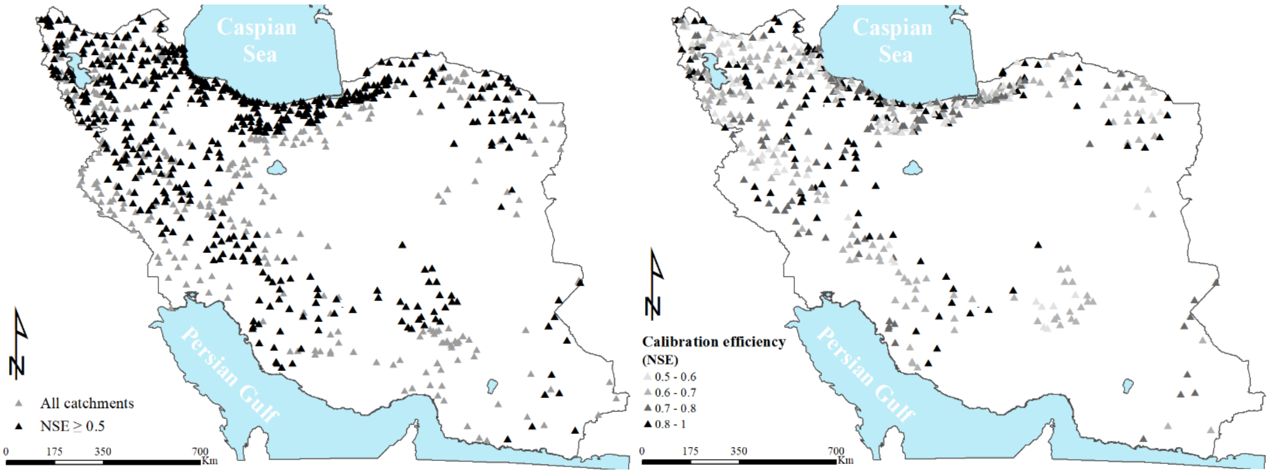

Fig. 3. Gauge location in Iran. Left panel: Location of all 963 catchments considered within Iran; the 576 catchments maintained (NSE ≥ 0.5) for parameter transfer strategy are shown by black triangles, the excluded catchments by grey triangles. Right panel: Average calibration efficiency of the two calibration periods over the 576 catchments. The triangles have been plotted at the outlet of catchments.

To compare the calibration results of the two periods, we plotted the NSE values obtained with HBV across the two calibration periods in 576 catchments (Fig. 4). The relationship between NSE values of these two calibration periods is somewhat weak (Pearson's r = 0.55), with fairly widespread data points scattered along with both sides of line 1:1. This indicates that model performance can vary quite a bit with the calibration period. For all 576 catchments, about 82% of catchments have an NSE value of more than 0.6 for two calibration periods. Figure 4 also demonstrates that humid (median NSE = 0.72) and semi-humid (median NSE = 0.66) catchments show in general better performance than the semi-arid (median NSE = 0.59) and arid (median NSE = 0.61) classes.

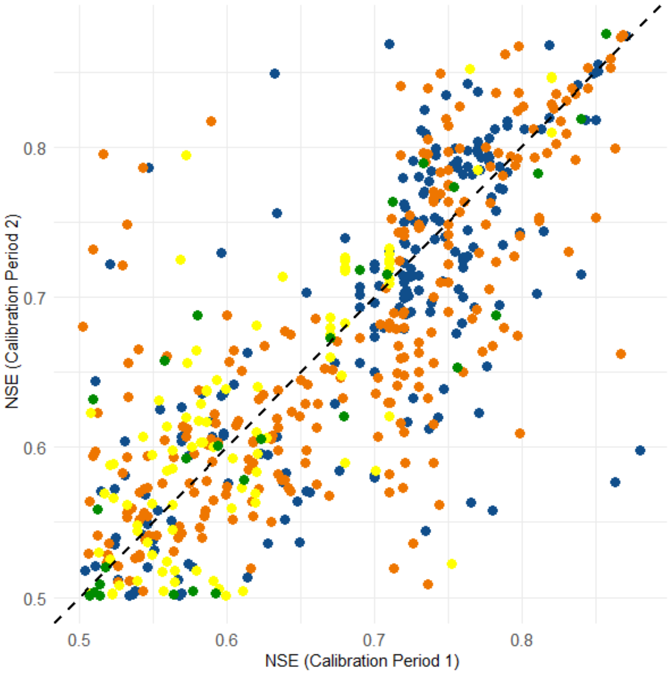

Fig. 4. Comparison of the NSE values between two calibration periods for all four climate classes (n = 576). Legend of the climate classes according to Fig. 2.

4.2. Parameter uncertainty in regionalization approaches

Here, we judge the model parameters in two ways. First, we assess the stability of the model parameters over time by comparing parameter distribution for different calibration periods. Second, we assess the difference between the calibrated parameter value and the regionalized parameter value obtained with spatial mode. As such, we assess parameter stability across space. Third, we analyze the impact of the difference in temperature (ΔT) and precipitation (ΔP) between two periods on calibrated model parameters. The insights obtained in this section can be used to understand and explain the regionalization results in the next section.

4.2.1. Stability of model parameters over time

Figure 5 shows the 1:1 comparison of 15 HBV model parameters (differentiated for four climate classes) during two calibration periods. The r_1 correlation coefficients and median values of calibrated parameters for two periods are presented in Table 4. The degree-day factor (DDF) shows most stability over time, as indicated by a correlation coefficient of 0.6. The weakest relationship is obtained for the storage coefficient for slow response, K2, with a correlation coefficient of 0.29. This confirms that not only model performance, as shown in section 4.2, but also parameter values, can vary quite a bit across calibration periods. The parameters of the model are less stable for semi-arid and arid classes compared to humid and semi-humid classes.

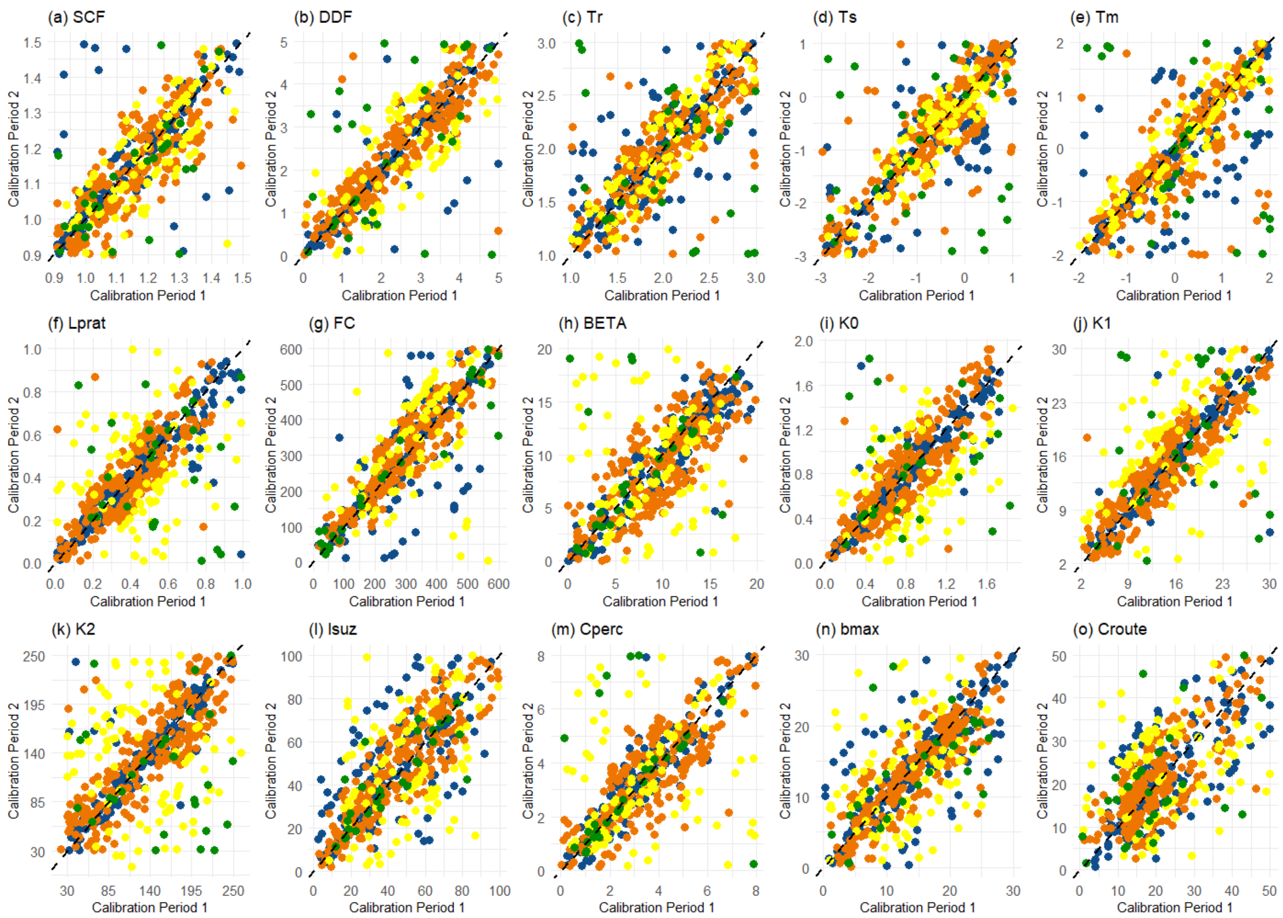

Fig. 5. 1:1 plot of all 15 HBV model parameter values for calibration periods. Legend of the climate classes according to Fig. 2. Legend of the climate classes according to Fig. 2.

To explore to what extent the variation in parameters during the two calibration periods can be attributed to a difference in temperature and precipitation in these two periods, we plotted the change in parameter value against ΔP and Delta ΔT (Fig. 6). It was found that among the 15 calibrated parameters (Fig. 5), eight parameters (SCF, DDF, Lprat, K1, K2, FC, BETA, and Cperc) seem to have remarkable variations between the two calibration periods.

Figure 6 shows that, at the country level, parameter values show an increasing trend parallel with an increase in ΔP except for SCF, K2, BETA, and Cperc (Fig. 6a to 6h). For ΔT, the difference in all parameters shows an increasing trend parallel with an increase of ΔT except for DDF, Lprat, FC, and BETA.

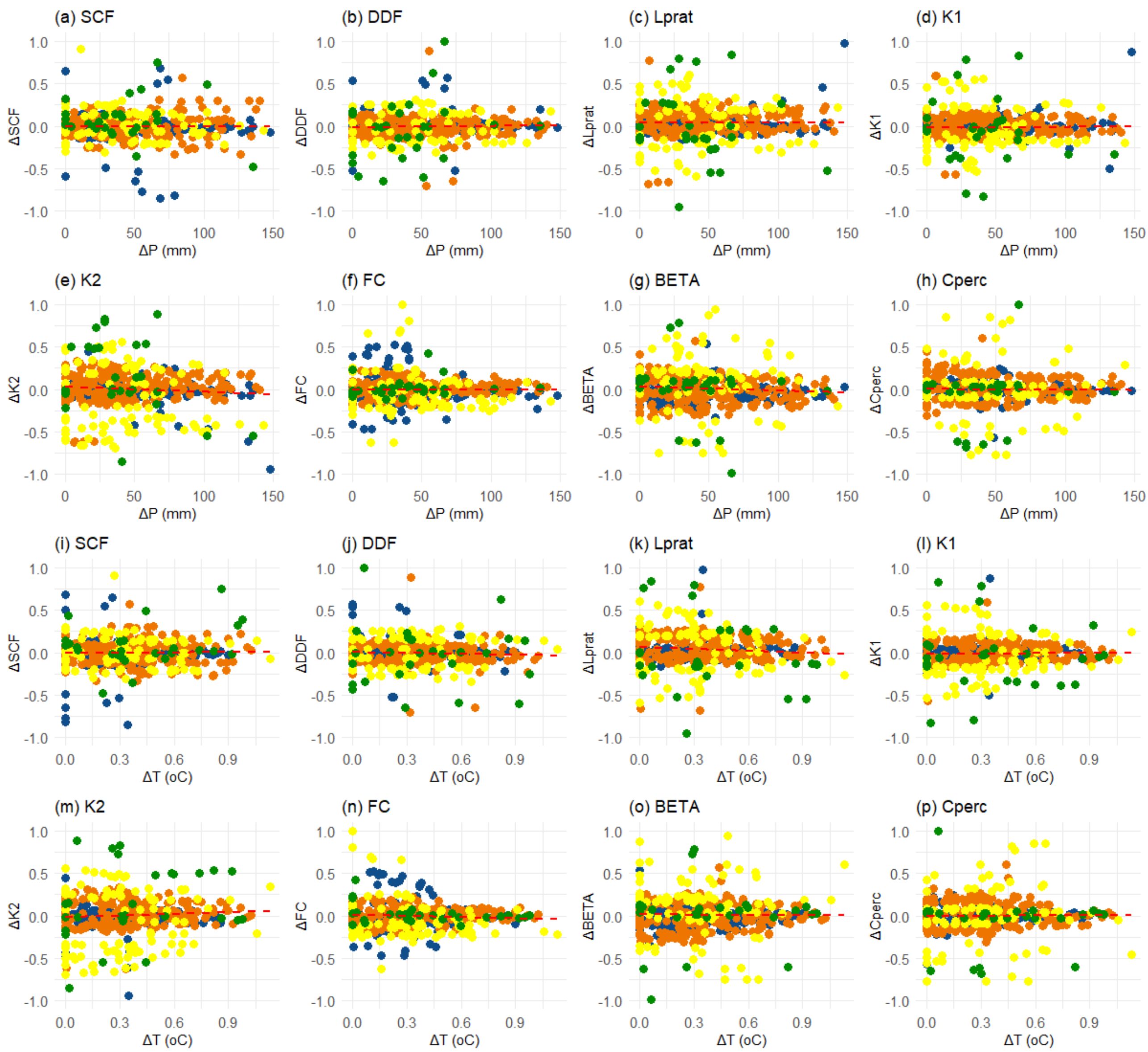

Fig. 6. 1:1 plot of difference in model parameters (ΔParameter) and (a to h) difference in precipitation (ΔP), (i to p) difference in temperature (ΔT). Legend of the climate classes according to Fig. 2.

4.2.2. Stability of parameters over space using regionalization

Here, we evaluate the difference between the original parameter value (obtained through calibration) and the regionalized parameter value using the best spatial mode; for each of the parameters, we evaluate the distance between both values for each catchment. Figure 7 shows the normalized parameter range for all 576 catchments. According to this distance distribution, the most and least robust parameters are DDF (r_2 = 0.61) and Croute (r_2 = 0.37) respectively. The r_2 correlation coefficients between calibrated and regionalized parameters are presented in Table 4.

Table 4

The calibrated parameters of the HBV model, the parameters value range (lower and upper limits), the correlation coefficient (r_1), median values of parameters for two calibration periods, and the correlation coefficient (r_2) between calibrated and regionalized parameters.

Parameter	Description	Lower	Upper	Median	r_1	r_2
SCF	Snow correction factor [-]	1	1.5	1.15	0.56	0.59
DDF	Degree day factor [mm/ºC day]	0	5	2.29	0.6	0.61
Tr	Threshold temperature above which precipitation is rain [ºC]	1	3	1.98	0.49	0.53
Ts	Threshold temperature below which precipitation is snow [ºC]	-3	1.0	-0.68	0.55	0.58
Tm	Threshold temperature above which melt starts [ºC]	-2	3	0.05	0.42	0.5
Lprat	Parameter related to the limit for potential evaporation [-]	0	1	0.42	0.42	0.39
FC	Field capacity [mm]	0	600	300.84	0.58	0.6
BETA	The nonlinear parameter for runoff production [-]	0	20	9.52	0.4	0.39
K0	Storage coefficient for very fast response [day]	0	2	0.8	0.46	0.55
K1	Storage coefficient for fast response [day]	2	30	15.13	0.51	0.45
K2	Storage coefficient for slow response [day]	30	180	135.68	0.29	0.41
lsuz	Threshold storage state	1	100	45.12	0.5	0.47
Cperc	Constant percolation rate [mm/day]	0	8	3.32	0.36	0.51
bmax	Maximum base at low flows [day]	0	30	13.79	0.44	0.45
Croute	Free scaling parameter [day^2 /mm]	0	50	20.61	0.31	0.37

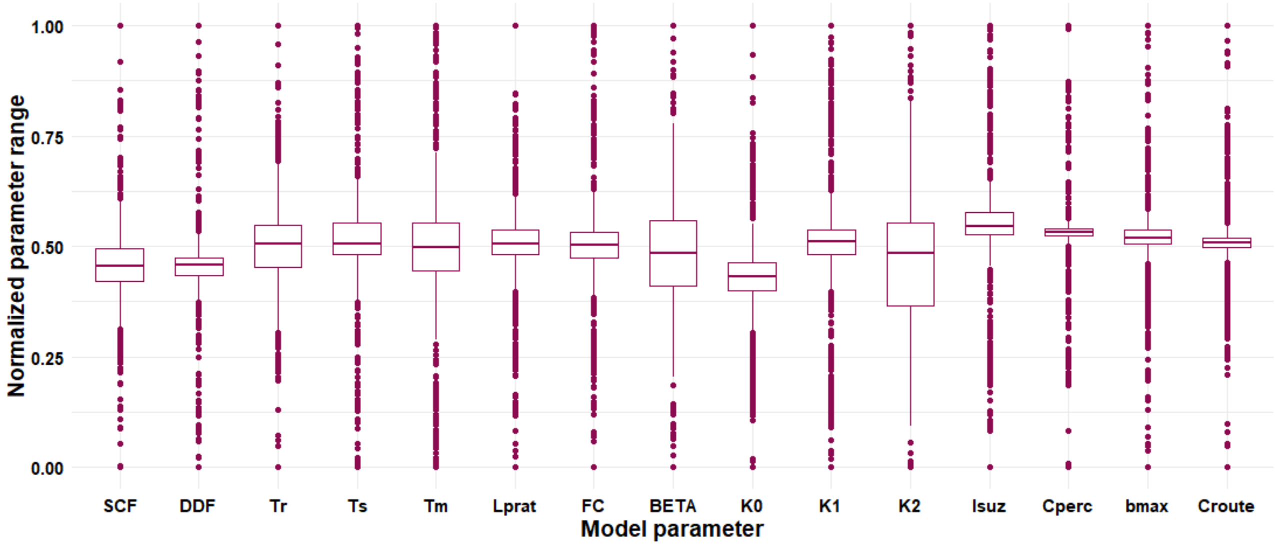

Fig. 7. Distance between calibrated and spatial mode regionalized parameter value, normalized over the parameter range of. Each boxplot contains 576 values; one for each catchment.

4.3. Model performance achieved with the different regionalization approaches

In the next step, we compare the performance of the HBV model between the two parameter transfer strategies and temporal mode. Figure 8 shows the boxplot comparison of NSE values for all four cases: calibration, temporal (TEM), spatial (SPA), and spatiotemporal (SPA_TEM). The best performance is for calibration mode (NSE = 0.75), followed by TEM (NSE= 0.64, decline of 15% compared to calibration), SPA (NSE = 0.46, decline of 39% and for SPA_TEM (NSE = 0.39, decline of 48%). The results from Fig. 8 show that the temporal regionalization performed better than the SPA and SPA_TEM.

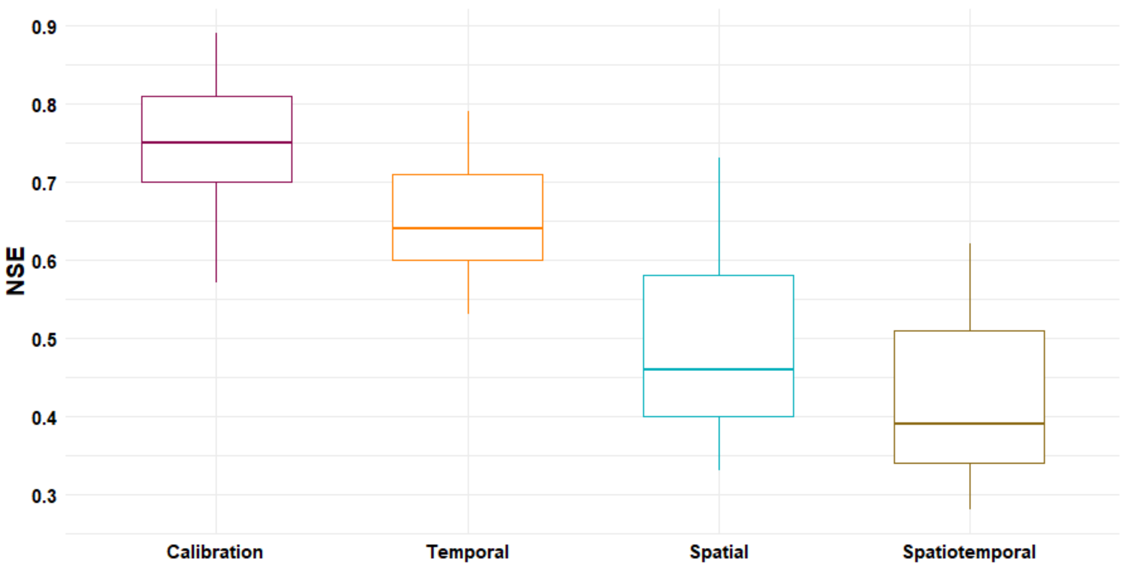

Fig. 8. Boxplot of the NSE values for calibration, temporal mode, and two tested strategies.

4.4. Accounting for temporal and spatial proximities

To further analysis the advantage that the temporal mode has over the other two (spatial and spatiotemporal) modes, we considered the following two scenarios:

Scenario 1: The spatial distribution is not suitable for some catchments, which results in a large distance between them and their nearest neighbor (donor) catchment. This scenario considers the SPA and SPA_TEM strategies at a clear disadvantage compared to TEM mode.

Scenario 2: The calibration periods 1 and 2 are consecutive and there is no temporal lag between them. If the meteorological inputs for some of catchments have not changed significantly over these two periods (especially in adjacent catchments), the temporal transfer mode may be superior to the other two strategies (Patil and Stieglitz, 2015).

To reduce the effects of the two considered scenarios, we repeated the parameter transfer strategies under individual conditions 1 and 2.

Individual Condition 1: We remove the catchments that have the nearest neighbor catchment more than 68 km away (the median nearest neighbor distance is 69.4 km). This reduces the number of catchments from 576 to 289.

Individual Condition 2: We consider a temporal lag distance (10 years) between two calibration periods so that calibration periods 1 changed to WY from 1993 to 1999 (instead of 1993-2001) and calibration period 2 changed to WY from 2009 to 2014 (instead of 2001-2008). Nonetheless, unlike individual condition 1, all 576 catchments are retained for simulations. It worth nothing that Merz et al. (2009) suggested that the minimum calibration period for interpreting the temporal variability of hydrological processes of a catchment is five years.

Figures. 9a and 9b show the boxplot of NSE values from four scenarios for individual conditions 1 and 2, respectively. The results for both cases are similar to those results for calibration and parameter transfer strategies (Fig. 8). For individual condition 1 the highest NSE value is obtained for the calibration scenario (NSE = 0.78), after that TEM (NSE = 0.67; decline of 13.7%), SPA (NSE = 0.51; decline of 34.5%) and then SPA_TEM (NSE = 0.43; decline of 44.6%) strategies. For individual condition 2, the median NSE values for calibration, TEM, SPA and SPA_TEM PTSs are: 0.73, 0.59 (decline of 18.9%), 0.44 (decline of 39.4%) and 0.36 (decline of 50.2%) respectively.

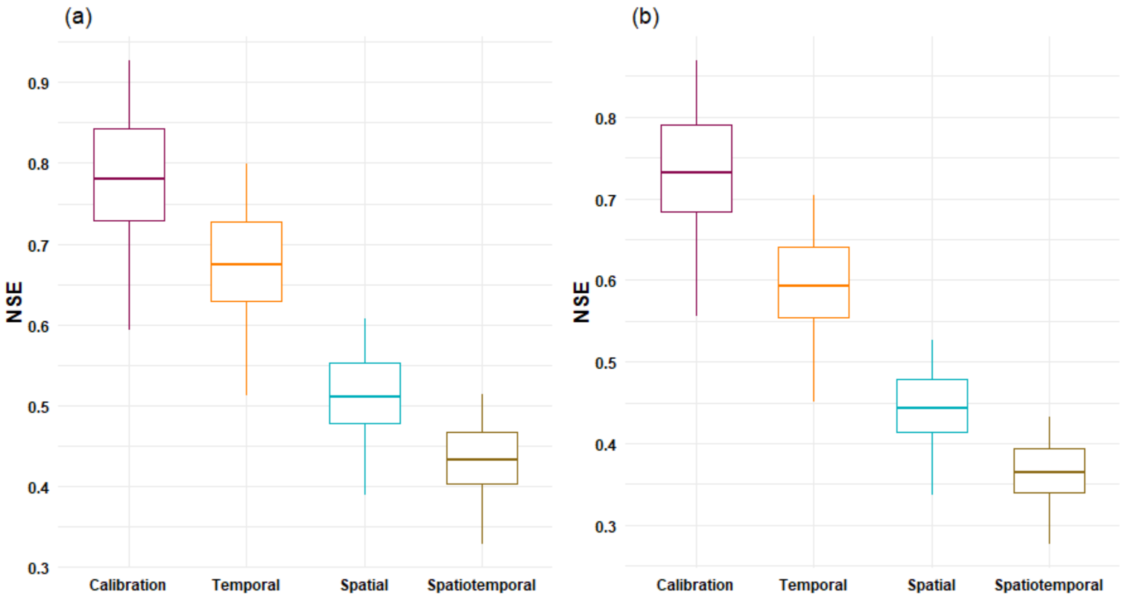

Fig. 9. Boxplot of the NSE values for calibration, temporal mode, and two tested strategies. (a) individual condition 1 and (b) individual condition 2.

4.5. Accounting for spatial distance to donor catchment

Here, we plot the NSE-calibration versus the (i) NSE-mode and (ii) delta-NSE, as well as NSE-decline versus distance to the donor catchment. Figure 10a shows the relationship between the calibration efficiency of the donor catchment and the model's efficiency on the ungauged catchment under three modes. Results suggest that using a well-modeled catchment as donor warrants a good level of efficiency of the mode for the ungauged catchment. However, conversely, if the poorly calibrated catchments are used as donors, the performances of the parameter transfer strategies are clearly affected. This effect is consistent across the three evaluated modes.

Figure 10b shows a decline in NSE value (ΔNSE) versus the distance to donor catchment for spatial and spatiotemporal. The greater distance leads to poorer performance of parameter transfer. The correlation coefficient between NSE-decline and the spatial distance is $r = 0.15$ and $r = 0.19$ for spatial and spatiotemporal, respectively. As seen in this Fig., there is an upward trend between NSE-decline and the spatial distance for both spatial and spatiotemporal. The use of linear regression quantifies that the effect of distance is somewhat stronger for spatiotemporal (R^2 = 0.039) compared to spatial (R^2 = 0.022) (Fig. 10b).

We consider 16 separate temporal lag distances between two calibration periods, so that the calibration periods 1 changed to 16 WYs from 1993 to 2014 (i.e., 1993-1999, 1994-2000, 1995-2001, 1996-2002, 1997-2003, 1998-2004, 1999-2005, 2000-2006, 2001-2007, 2002-2008, 2003-2009, 2004-2010, 2005-2011, 2006-2012, 2007-2013, and 2008-2014) and calibration period 2 changed to WY from 2009 to 2014 (instead of 2001-2008). Figure 10c shows the NSE for temporal and spatiotemporal (y-axis) against NSEs for each WY (x-axis). As seen in Fig. 10c, reducing the temporal lag between calibration periods reduces the difference between their temporal/spatiotemporal performance. This is a gradual decline in performance so that the greatest difference is between WY 1993-1999 and calibration period 2 (2009-2014) for temporal, but the WY changes to 1995-2001 and 1996-2002 for spatiotemporal. The linear regression for temporal distance quantifies that the effect of temporal distances is much stronger for spatiotemporal (Fig. 10c) than the spatiotemporal under spatial distance (Fig. 10b).

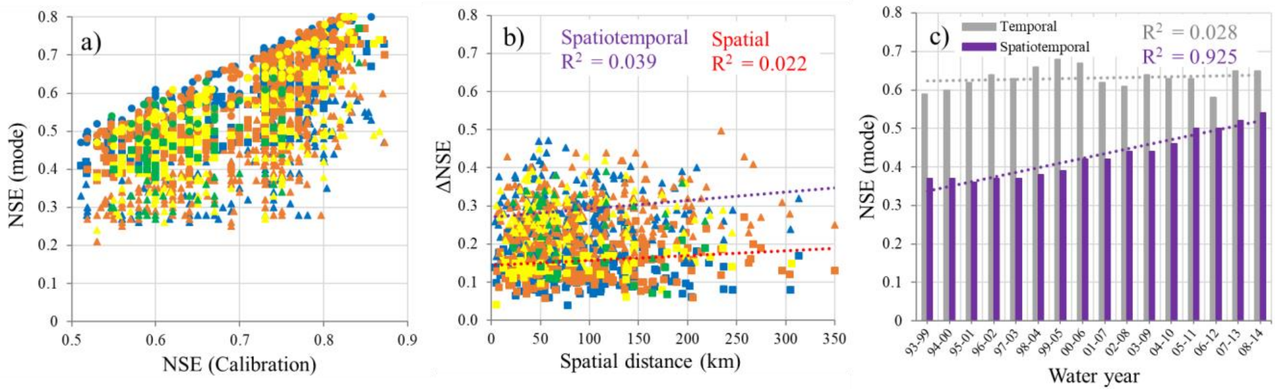

Fig. 10. (a) NSE-calibration versus NSE-mode; (b) NSE-decline (ΔNSE) versus spatial distance from donor catchment for spatial and spatiotemporal and (c) NSE-mode versus the water years between both periods. Dotted lines indicate the trend lines. Circle, square, and triangle indicate temporal, spatial, and spatiotemporal, respectively. Legend of the climate classes follows Fig. 2.

4.6. Controls on model transferability

Figure 11 shows the correlation matrix between catchment characteristics and parameter transfer strategies. As seen in this Fig., dynamic characteristics have a relatively strong relationship with ΔNSE for local and regional scales. The main reason is that precipitation, temperature, and snow cover changes between the two periods have a remarkable impact on model transferability (Table 3). Among these dynamic characteristics, aridity has the strongest correlation with ΔNSE for both local and regional scales. This correlation is more significant for the local scale than the regional one. ΔP has a more complicated correlation with ΔNSE between two scales. For all three parameter transfer strategies, its correlation is positive in arid and semi-arid catchments, whereas it is negative in the humid, semi-humid catchments, and when evaluated at the country level. The correlation between ΔT and ΔNSE is negative for humid, semi-humid, semi-arid, and Iran, whereas it is positive for the arid class. Δsnow shows a negative correlation with ΔNSE for both regional and local scales. Its correlation is stronger at the regional scale than the local one.

In the case of the static characteristics, there is a positive correlation between area and ΔNSE for semi-humid, semi-arid, arid, and Iran, whereas it is negative for humid. The whole country has a greater correlation than the regional scale. For the elevation case, this correlation is negative for two regional and local scales. The elevation has a more significant impact on model transferability than the catchment area for both scales.

In general, correlation coefficients are higher for the temporal transfer strategy than for the spatial and spatiotemporal strategies, aridity and elevation are the two main controls on model transferability, both at the regional (climate classes) and national scale.

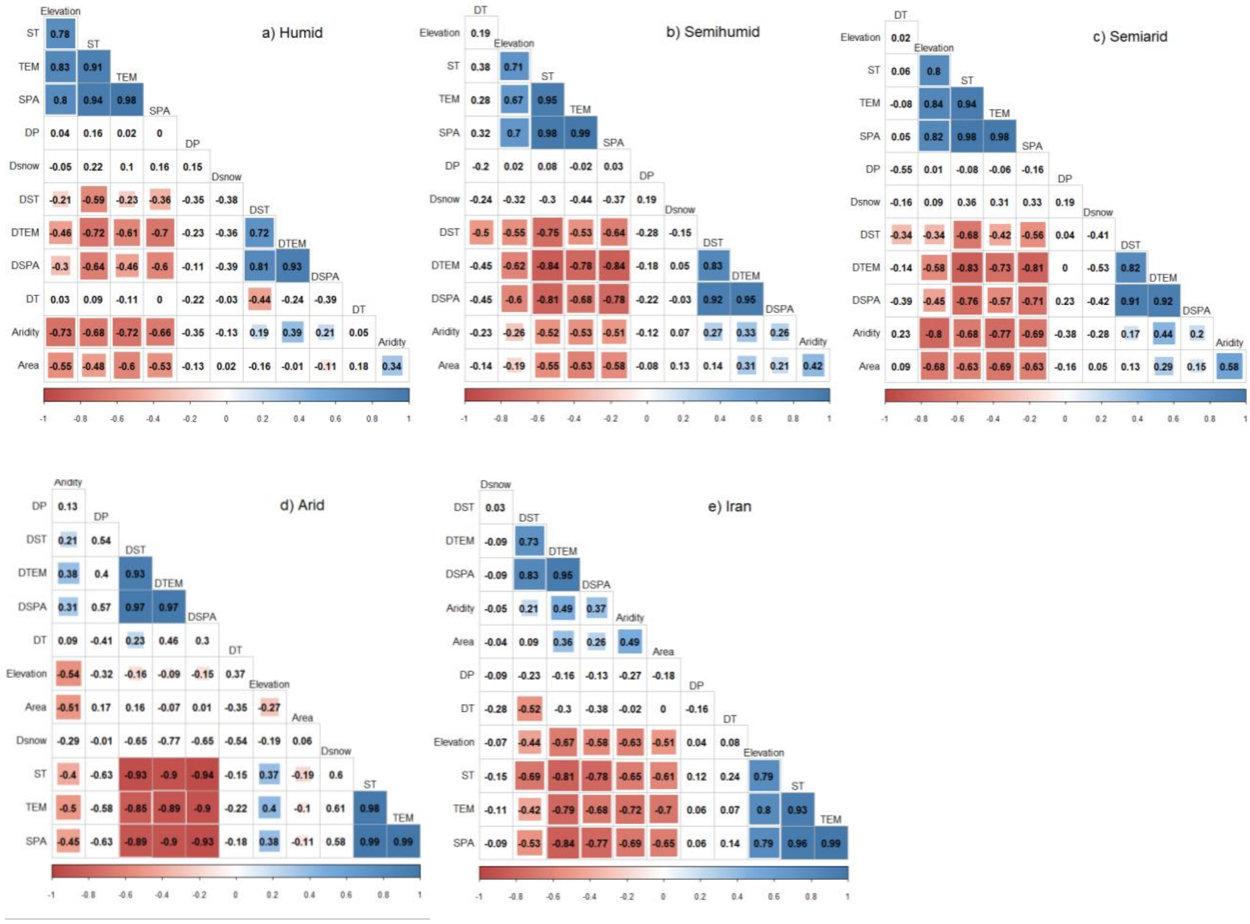

Fig. 11. Correlation matrix between catchment characteristics and parameter transfer strategies. TEM is temporal, SPA is spatial, ST is spatiotemporal, DP is ΔP, DT is ΔT, Dsnow is Δsnow, DTEM, DSPA, and DST are ΔNSE for temporal, spatial, and spatiotemporal, respectively. The values are correlation coefficients.

5. Discussion

Comparing the model performance in terms of NSE values (Fig. 4) and model parameters (Fig. 5) by applying two calibration periods demonstrated that the performance of the HBV model did not change significantly for the same catchments between two different periods. Median, minimum, and maximum ΔNSE values between two calibration periods are 0.001, 0, and 0.28, respectively, between 576 study catchments. Similar conclusions have been demonstrated by Vaze et al. (2010), Razavi and Tolson (2013), and Patil and Stieglitz (2015). Dispersion of data points (median NSE values for each catchment) in Figs. 4 and 5 indicates a lack of systematic bias, which clarifies that there is no superiority of one calibration period over the other (calibration period 1 against 2). About 69.8% of our catchments (402 out of 576), the difference between optimal NSEs for calibration periods 1 and 2 is less than 10% (median NSE = 6.8%). This is consistent with the conclusion by Merz et al. (2011), who demonstrated that NSE for HBV model, across 273 catchments in Austria, showed small variability across six consecutive five years for calibration. This finding is also consistent with Patil and Stieglitz (2015), who demonstrated that KGE values (Gupta et al., 2009) for the EXP-HYDRO model (Patil and Stieglitz, 2014; Patil et al., 2014b, 2014a) across 294 catchments in the U.S., showed small variability across two calibration periods. Although, the difference in performance between the two calibration periods is highly dependent on (i) the length of the calibration period and (ii) climate variability within the calibration period (Yapo et al., 1996). Variation in climate between the two periods is small (Table 3), but still can lead to reduction in model performance from calibration to transferability, especially at the local scale. Note that the catchments used in the study were already selected to have a performance of at least 0.5 in both calibration periods, thereby already limiting the variability between both periods. As noted in Section 4.1, the model performed better in humid catchments compared to dry catchments. The origin of this model is in humid regions (Sweden), so that it seems to perform better in humid regions or arid regions with periodic climatic conditions similar to humid regions.

The results of calibrated parameters (Fig. 5) show that there is temporal variability in the 15 calibrated parameter values. This variability can be considered as "uncertainty", as the parameter values are calibration period dependent. According to Fig. 4, the parameters with the largest and smallest scatter show the highest and lowest uncertainties, respectively. Therefore, the most and least uncertain parameters are K2 (r_1 = 0.29) and DDF (r_1 = 0.6) respectively. Interestingly, in Seibert (1997), Uhlenbrook et al. (1999), Merz and Blöschl (2004), the most uncertain parameters were Cperc, FC and K2, respectively, and the least uncertain ones were DDF, DDF, and K1 respectively.

Our results showed that Cperc (r_1 = 0.36), Croute (r_1 = 0.31) and K2 (r_1 = 0.29) were known as the three most sensitive parameters to the changes across the calibration periods. In a more detailed investigation, the uncertainty of the parameters is investigated by calculating the parameter distance distribution. The results show that DDF, FC, SCF, and Ts have lower uncertainty, whereas Croute, BETA, Lprat, and K2 have higher uncertainty. One of the reasons for the high uncertainty of Croute is that its value tends to be large in lowland and mountainous catchments (about 12%), implying a more non-linear channel response in these catchments (increasing discharge result in faster response), but the hydrological reason for these patterns is unclear. This pattern was shown in Merz and Blöschl (2004) for some of the catchments in northern Austria. Two other parameters (Cperc and K2) are affected by the runoff generation conditions of the catchments, not the input data conditions during two calibration periods. The most stable parameters in terms of temporal are DDF (r_1 = 0.6), FC (r_1 = 0.58) and SCF (r_1 = 0.56) which represent: (1) degree day factor, (2) field capacity and (3) snow correction factor which are unlikely to be affected by severe temporal changes during calibration periods. These three parameters are affected by snow and soil conditions of calibration periods and studied catchments. Generally, the parameters related to snow and runoff routines are the lowest and highest uncertain parameters compared to the other model routines.

Nevertheless, the parameters, which are stables in terms of temporal variability, are different for different types of the rainfall-runoff model used. Patil and Stieglitz (2015) calibrated EXP-HYDRO in 294 catchments in the U.S. for two calibration periods. They concluded that the TEM variability is different between all parameters. In their study, the parameter *f*, showed the lowest uncertainty (highest correlation) between calibration periods (lowest uncertainty). Merz and Blöschl (2004) calibrated the HBV model in 308 catchments of Austria for two calibration periods and founded that the range of R^2 between two sets of calibrated parameters ranged between 0.09 and 0.64 (only 5 out of 11 parameters have $R^2 \geq 0.5$). Oudin et al. (2008) compared TOPMO (6 parameters) and GR4J (4 parameters) models for 913 catchments in France. Their results showed the superiority of the GR4J model compared to TOPMO in terms of a higher correlation between the parameters across calibration periods. Although, it is difficult to judge how the model parameters values (for all hydrological rainfall-runoff models used in the literature on the model parameters transfer) will change in response to wet and dry periods and land-use changes within catchments. This insight requires more extensive and comprehensive research (Eckhardt et al., 2003; Patil and Stieglitz, 2015; Wang and Kalin, 2011). Thus, even when considering variable parameters (Section 4.2.1), model transferability will still be prone to error. The error became larger when study catchments behaved differently under extreme events. Other possible effects of climate variability in the long-term are changes in vegetation and the water table that need to be assessed in data-sparse catchments of Iran.

Our results from the parameter transfer strategies (Fig. 8) show the overall superiority of the temporal (TEM) transfer mode over spatial (SPA) and spatiotemporal (SPA_TEM) transfer strategies. Individual comparison of them across all 576 catchments clarified that the TEM mode has the best performance at 414 catchments (and worst at 88 catchments), the SPA strategy is best in 118 catchments (and worst at 219 catchments), and the SPA_TEM strategy is best in 44 catchments (and worst at 269 catchments). Figure 12 shows the location of catchments, where either the SPA or SPA_TEM is the best case. No specific and regular geographic pattern is deduced from the spatial distribution of tested catchments in terms of the superiority of SPA and SPA_TEM strategies over the TEM case. Table 5 shows the comparison of two groups of transfer strategies between studied catchments (Group 1: TEM mode performing best; Group 2: SPA or SPA_TEM strategy performing best) in terms of three hydroclimatic indices: aridity index (PET/P), annual runoff ratio (Q/P) and mean annual rainfall (P). Even though the median values of these three indices show that the catchments in group 1 are wetter (lower and higher values of PET/P and P, respectively) and less flashy (lower Q/P). Nevertheless, Fig. 12 shows in some parts of the study area with a low density of catchments and larger distance between neighboring catchments (e.g., in the central, southwestern and southeastern parts of Iran), the TEM mode is not superior to SPA and SPA_TEM strategies. This relatively irregular geographical pattern was also demonstrated in the U.S. by Patil and Stieglitz (2015), who found that in regions with low catchment density, the temporal mode of parameter transfer does not always outperform the other two strategies.

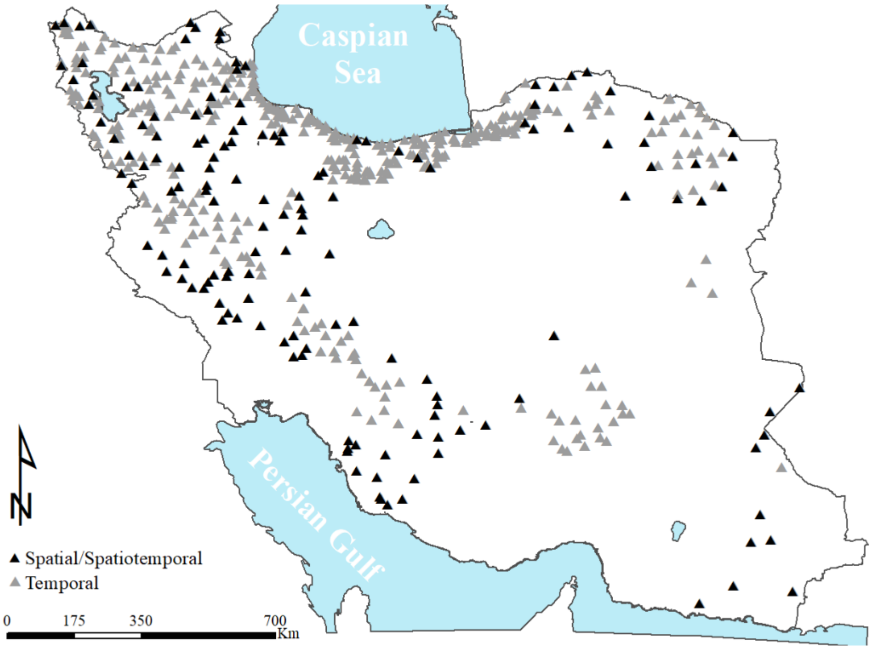

Fig. 12. Location of catchments where the temporal mode performs best (grey triangles) and catchments where either spatial or spatiotemporal perform best (black triangles). The triangles have been plotted at the outlet of catchments.

Table 5

Median values of three hydro-climatic descriptors for two catchment groups (shown by Fig. 12) (n = 576). Numbers in parentheses are standard deviation values.

Group No.	Best performance	*P* (mm)	Q/P	PET/P
1	Temporal	774 (245)	0.38 (0.09)	0.59 (0.33)
2	Spatial or spatiotemporal	387 (188)	0.42 (0.12)	1.14 (0.38)

The evaluation of characteristics (aridity, differences in precipitation, temperature and snow cover of calibration and validation periods, catchment area and elevation) on model transferability showed that the slight climatic variability between the two calibration periods has a remarkable effect on model transferability. Our result shows that aridity and catchment elevation are the two major controls on model transferability at two regional (climate classes) and local (the whole country) scales. These effects are most robust at a regional scale. This finding shows that negligible climatic variability affects the model transferability more at a smaller scale than a larger scale. The more noticeable effect of static characteristics (catchment area and elevation) on model transferability indicates that these two physical characteristics are useful descriptors for model transferability.

Under the two individual conditions, model transferability shows that the TEM mode retained its superiority over the SPA and SPA_TEM strategies. For *Individual Condition* 1, when only those catchments with the nearest neighbor < 68 km away are maintained, the SPA and SPA_TEM strategies showed performance improvement compared to the base scenario, so that their differences with calibration is about 3.59% lower than the base scenario (in terms of NSE value). This is an expected result because reducing the spatial distance between donor and target catchments will most likely reduce the spatial variability of hydrological behavior and improve performance (Oudin et al., 2008; Patil and Stieglitz, 2015). This finding is confirmed by testing the relationship between ΔNSE and spatial distance (Fig. 10b) so that increasing spatial distance between donor and target catchments increases ΔNSE for SPA and SPA_TEM (see section 4.5). For *Individual Condition* 2 (when a 10-year temporal gap is added between two calibration periods), the NSE difference between calibration and the TEM mode is 4.51% higher than the base scenario and is virtually unchanged between calibration and two other strategies.

In a more accurate evaluation, linear regression confirmed that as the temporal distance between calibration periods increases, the performance of the SPA_TEM strategy decreases with a greater slope compared to the increased spatial distance between the donor and target catchments ($R^2 = 0.925$ vs. $R^2 = 0.039$). Therefore, an increase in the temporal lag between calibration and validation periods reduces the model performance gap between the TEM mode and the SPA and SPA_TEM strategies more than the case with an increase in spatial distance.

Overall, our finding is consistent with Patil and Stieglitz (2015) across 294 catchments in the U.S. Here, the temporal gap between calibration and TEM mode as well as between calibration and the

SPA and SPA_TEM strategies is 10 years. Maybe one reason for this higher difference between our results with Patil and Stieglitz (2015) is due to a longer temporal gap (10-year versus 8-year). Nevertheless, this 10-year temporal gap reduced the performance of the examined strategies, but it is not entirely clear how these strategies will compare for larger (i.e., >30 years or more) temporal gaps.

5. Conclusion

This study, which is so far the most comprehensive PUB study in Iran, investigated three strategies for transferring model parameters, including temporal, spatial, and spatiotemporal, during two calibration periods using a conceptual rainfall-runoff model called HBV at 576 unregulated catchments across Iran. Our results showed that temporal mode has the best performance, with the lowest decline in performance than calibration (median decline of 14.66%), while these declines for spatial and spatiotemporal were 38.6% and 48%, respectively. Thus, we conclude that the stability of the parameters of the HBV model in the temporal mode is higher than for spatial strategy. Hence, temporal comes out best, making sense because the model has "seen" the catchment. This finding is in accordance with previous studies by Zhang and Chiew (2009) in Australia, Parajka et al. (2005) in Austria, and Patil and Stieglitz (2015) in the U.S. We also showed that the superiority of temporal mode is maintained under three scenarios ((1) a decrease in the spatial (geographical) distance between donor and target catchments; and (2) an increase in the temporal lag (10 years) between calibration and validation periods and (3) a gradual increase in the temporal lag between calibration and validation periods). This finding is consistent with a previous study by Patil and Stieglitz (2015), but with relatively poorer performance at both temporal and spatial transfer strategies. We also concluded that an increase in temporal lag between calibration and validations leads to a reduction in the model performance gap between the temporal mode and spatial and spatiotemporal strategies. This finding suggests that spatiotemporal parameter transfer can be a fairly reliable option for PUB studies and climate change-related studies, at least in wetter catchments. Our results are obtained from two consecutive calibration periods (by averaging model parameters). However, more research is needed to: (i) examine longer calibration periods, (ii) obtain a stronger relationship between temporal mode and spatial strategy, (iii) explore the effects of increasing time lag (>10 years) between calibration and validation periods on results to achieve more confident about the variability of the hydrological model parameters. Exploring dynamic (aridity and differences in precipitation, temperature and snow cover of calibration and validation periods) and static (catchment area and mean elevation) controls on model transferability showed that aridity and catchment elevation are two major controls on model transferability at two regional (climate classes) and local (the whole country) scales.

Finally, it can be inferred that if there is a catchment for which only a limited period of observations is available, it is preferred to calibrate the model on this limited period above having parameters from a donor catchment. However, in many cases, even a limited period of observations will not be available, and spatial or spatiotemporal are the only options left.

Acknowledgments

The authors are grateful to Iran Energy Ministry (IEM), Iran Meteorological Organization (IMO), Iran National Cartographic Center (INCC), and Iran Natural Resources Organization (INRO) for their cooperation and for providing the required data. We would like to thank Remko Uijlenhoet for supporting this study.

References

Ardia, D., Boudt, K., Carl, P., Mullen, K.M., Peterson, B.G., 2011. Differential evolution with deoptim. R J. 3, 27–34. https://doi.org/10.32614/rj-2011-005

Arsenault, R., Brissette, F.P., 2014. Continuous streamflow prediction in ungauged basins: The effects of equifinality and parameter set selection on uncertainty in regionalization approaches. Water Resour. Res. 50, 6135–6153. https://doi.org/10.1002/2013WR014898

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. 466–467, 37–46. https://doi.org/10.1016/j.jhydrol.2012.07.048

Bárdossy, A., 2007. Calibration of hydrological model parameters for ungauged catchments. Hydrol. Earth Syst. Sci. 11, 703–710. https://doi.org/10.5194/hess-11-703-2007

Bergström, S., 1976. Development and application of a conceptual runoff model for Scandinavian catchments. Series A, Lund Institute of Technology,. University of Lund, Lund, Sweden. No. 52.

Blöschl, G., Sivapalan, M., Wagener, M., Viglione, A., Savenije, H., 2013. Runoff prediction in ungauged basins: Synthesis across processes, places and scales, Cambridge University Press, Cambridge, U.K. https://doi.org/10.1002/2014EO020025

Brigode, P., Oudin, L., Perrin, C., 2013. Hydrological model parameter instability: A source of additional uncertainty in estimating the hydrological impacts of climate change? J. Hydrol. 476, 410–425. https://doi.org/10.1016/j.jhydrol.2012.11.012

Chiew, F.H.S., Vaze, J., Viney, N., Jordan, P., Perraud, J., Zhang, L., Teng, J., Young, W., Arancibia, J.P., Morden, R.A., Freebairn, A., Austin, J., Hill, P., Wiesenfield, C., Murphy., R., 2008. Rainfall-runoff modelling across the Murray-Darling Basin. Rainfall-runoff modelling across the Murray-Darling Basin. A Report to the Australian Government from the CSIRO Murray-Darling Basin Sustainable Yields Project. CSIRO.

Choubin, B., Solaimani, K., Rezanezhad, F., Habibnejad Roshan, M., Malekian, A., Shamshirband, S., 2019. Streamflow regionalization using a similarity approach in ungauged basins: Application of the geo-environmental signatures in the Karkheh River Basin, Iran. Catena 182, 104128. https://doi.org/10.1016/j.catena.2019.104128

Clark, G.E., Ahn, K.-H., Palmer, R.N., 2017. Assessing a Regression-Based Regionalization Approach to Ungauged Sites with Various Hydrologic Models in a Forested Catchment in the

Northeastern United States. J. Hydrol. Eng. 22, 05017027. https://doi.org/10.1061/(asce)he.1943-5584.0001582

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. https://doi.org/10.1016/j.jhydrol.2017.04.032

De Martonne E., 1926. Une nouvelle fonction climatologique: L'indice d'aridité. La Meteorologie, 2:449–458

Eckhardt, K., Breuer, L., Frede, H.G., 2003. Parameter uncertainty and the significance of simulated land use change effects. J. Hydrol. 273, 164–176. https://doi.org/10.1016/S0022-1694(02)00395-5

Gupta, H. V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. J. Hydrol. 377, 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003

Haimberger, L., 2007. Homogenization of radiosonde temperature time series using innovation statistics. J. Clim. 20, 1377–1403. https://doi.org/10.1175/JCLI4050.1

Hargreaves, G.L., Hargreaves, G.H., Riley, J.P., 1985. Irrigation Water Requirements for Senegal River Basin. J. Irrig. Drain. Eng. 111, 265–275. https://doi.org/10.1061/(asce)0733-9437(1985)111:3(265)

Hrachowitz, M., Savenije, H.H.G., Blöschl, G., McDonnell, J.J., Sivapalan, M., Pomeroy, J.W., Artheimer, B., Blume, T., Clark, M.P., Ehret, U., Fenicia, F., Freer, J.E., Gelfan, A., Gupta, H.V., Hughes, D.A., Hut, R.W., Montanari, A., Pande, S., Tetzlaff, Troch, P.A., Uhlenbrook, S., 2013. A decade of Predictions in Ungauged Basins (PUB)-a review. Hydrol. Sci. J. 58, 1198-1255. https://doi.org/10.1080/02626667.2013.803183

Iran Energy Ministry (IEM)., 2016. Comprehensive assessment of Iran's water resources measurement stations, Iran (in Persian).

Iran Energy Ministry (IEM)., 2018. Surface water resources dataset for Iran, Iran (in Persian).

Iran Meteorological Organization (IMO)., 2018. Temperature and precipitation dataset for Iran, Iran (in Persian).

Juston, J., Seibert, J., Johansson, P.-O., 2009. Temporal sampling strategies and uncertainty in calibrating a conceptual hydrological model for a small boreal catchment. Hydrol. Process. 23, 3093–3109. https://doi.org/10.1002/hyp.7421

Kay, A.L., Jones, D.A., Crooks, S.M., Kjeldsen, T.R., Fung, C.F., 2007. An investigation of site-similarity approaches to generalisation of a rainfall-runoff model. Hydrol. Earth Syst. Sci. 11, 500–515. https://doi.org/10.5194/hess-11-500-2007

Kennard, M.J., Pusey, B.J., Olden, J.D., MacKay, S.J., Stein, J.L., Marsh, N., 2010. Classification of natural flow regimes in Australia to support environmental flow management. Freshw. Biol. 55, 171–193. https://doi.org/10.1111/j.1365-2427.2009.02307.x

Lee, H., McIntyre, N., Wheater, H., Young, A., 2005. Selection of conceptual models for regionalisation of the rainfall-runoff relationship. J. Hydrol. 312, 125–147. https://doi.org/10.1016/j.jhydrol.2005.02.016

Lidén, R., Harlin, J., 2000. Analysis of conceptual rainfall-runoff modelling performance in different climates. J. Hydrol. 238, 231–247. https://doi.org/10.1016/S0022-1694(00)00330-9

Love, D., Uhlenbrook, S., Corzo-Perez, G., Twomlow, S., van der Zaag, P., 2010. Relations précipitation-interception-évaporation-écoulement dans un bassin versant semi-aride (nord du Limpopo, Zimbabwe). Hydrol. Sci. J. 55, 687–703. https://doi.org/10.1080/02626667.2010.494010

Masih, I., Maskey, S., Uhlenbrook, S., Smakhtin, V., 2011. Assessing the Impact of Areal Precipitation Input on Streamflow Simulations Using the SWAT Model. J. Am. Water Resour. Assoc. 47, 179–195. https://doi.org/10.1111/j.1752-1688.2010.00502.x

Masih, I., Uhlenbrook, S., Maskey, S., Ahmad, M.D., 2010. Regionalization of a conceptual rainfall-runoff model based on similarity of the flow duration curve: A case study from the semi-arid Karkheh basin, Iran. J. Hydrol. 391, 188–201. https://doi.org/10.1016/j.jhydrol.2010.07.018

McIntyre, N., Lee, H., Wheater, H., Young, A., Wagener, T., 2005. Ensemble predictions of runoff in ungauged catchments. Water Resour. Res. 41, 1–14. https://doi.org/10.1029/2005WR004289

Merz, R., Blöschl, G., 2004. Regionalisation of catchment model parameters. J. Hydrol. 287, 95–123. https://doi.org/10.1016/j.jhydrol.2003.09.028

Merz, R., Parajka, J., Blöschl, G., 2009. Scale effects in conceptual hydrological modeling. Water Resour. Res. 45 (9), W09405. http://dx.doi.org/10.1029/2009WR007872

Merz, R., Parajka, J., Blöschl, G., 2011. Time stability of catchment model parameters: Implications for climate impact analyses. Water Resour. Res. 47, 1–17. https://doi.org/10.1029/2010WR009505

Mitchell, M., 1998. An Introduction to Genetic Algorithms, MIT Press, 55 Hayward St., Cambridge, MA, United States.

Modallakdoust, S., Bayat, F., Soltani, B., Soleimani, K., 2008. Applying digital elevation model to interpolate precipitation. J. Appl. Sci. 8, 1471–1478. https://doi.org/10.3923/jas.2008.1471.1478

Nash, J.E., Sutcliffe, J. V., 1970. River flow forecasting through conceptual models part I - A discussion of principles. J. Hydrol. 10, 282–290. https://doi.org/10.1016/0022-1694(70)90255-6

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, 459–478. https://doi.org/10.2166/nh.2019.031

Oudin, L., Andréassian, V., Perrin, C., Michel, C., Le Moine, N., 2008. Spatial proximity, physical similarity, regression and ungaged catchments: A comparison of regionalization approaches based on 913 French catchments. Water Resour. Res. 44, 1–15. https://doi.org/10.1029/2007WR006240

Parajka, J., Blöschl, G., Merz, R., 2007. Regional calibration of catchment models: Potential for ungauged catchments. Water Resour. Res. 43. https://doi.org/10.1029/2006WR005271

Parajka, J., Merz, R., Blöschl, G., 2005. A comparison of regionalisation methods for catchment model parameters. Hydrol. Earth Syst. Sci. 9, 157–171. https://doi.org/10.5194/hess-9-157-2005

Parajka, J., Viglione, A., 2012. TUWmodel: Lumped hydrological model developed at the Vienna University of Technology for education purposes, R package version 0.1-2. http://CRAN.R-project.org /package=TUWmodel. (Jun. 21, 2014).

Parajka, J., Viglione, A., Rogger, M., Salinas, J.L., Sivapalan, M., Blöschl, G., 2013. Comparative assessment of predictions in ungauged basins - Part 1: Runoff-hydrograph studies. Hydrol. Earth Syst. Sci. 17, 1783-1795. https://doi.org/10.5194/hess-17-1783-2013

Patil, S., Stieglitz, M., 2014. Modelling daily streamflow at ungauged catchments: what information is necessary? Hydrol. Process. 28, 1159–1169. https://doi.org/10.1002/hyp.9660

Patil, S.D., Stieglitz, M., 2015. Comparing spatial and temporal transferability of hydrological model parameters. J. Hydrol. 525, 409–417. https://doi.org/10.1016/j.jhydrol.2015.04.003

Patil, S.D., Wigington, P.J., Leibowitz, S.G., Comeleo, R.L., 2014a. Use of hydrologic landscape classification to diagnose streamflow predictability in Oregon. J. Am. Water Resour. Assoc. 50, 762–776. https://doi.org/10.1111/jawr.12143

Patil, S.D., Wigington, P.J., Leibowitz, S.G., Sproles, E.A., Comeleo, R.L., 2014b. How does spatial variability of climate affect catchment streamflow predictions? J. Hydrol. 517, 135–145. https://doi.org/10.1016/j.jhydrol.2014.05.017

Perrin, C., Michel, C., Andréassian, V., 2001. Does a large number of parameters enhance model performance? Comparative assessment of common catchment model structures on 429 catchments. J. Hydrol. 242, 275–301. https://doi.org/10.1016/S0022-1694(00)00393-0

Petheram, C., Rustomij, P., Vleeshouwer, J., 2009. Rainfall-runoff modelling across northern Australia. A report to the Australian Government from the CSIRO Northern Australia Sustainable Yields project.

Pethveream, C., Bristow, K.., 2008. Towards an understanding of the hydrological factors, constraints and opportunities for irrigation in northern Australia: A review [WWW Document]. CSIRO L. Water Sci. Rep. URL https://publications.csiro.au/rpr/pub?pid=procite:99530409-31bc-4570-

97a6-83d715506c56 (accessed 3.31.21).

Pool, S., Viviroli, D., Seibert, J., 2017. Prediction of hydrographs and flow-duration curves in almost ungauged catchments: Which runoff measurements are most informative for model calibration? J. Hydrol. 554, 613–622. https://doi.org/10.1016/j.jhydrol.2017.09.037

Post, D.A., Jakeman, A.J., 1996. Relationships between catchment attributes and hydrological response characteristics in small Australian mountain ash catchments. Hydrol. Process. 10.

Rahimi, J., Ebrahimpour, M., Khalili, A., 2013. Spatial changes of extended de Martonne climatic zones affected by climate change in Iran. Theor. Appl. Climatol. 112, 409-418. https://doi.org/10.1007/s00704-012-0741-8

Razavi, S., Tolson, B.A., 2013. An efficient framework for hydrologic model calibration on long data periods. Water Resour. Res. 49 (12), 8418-8431. http://dx.doi.org/10.1002/2012WR013442

Refsgaard, J.C., Knudsen, J., 1996. Operational Validation and Intercomparison of Different Types of Hydrological Models. Water Resour. Res. 32, 2189–2202. https://doi.org/10.1029/96WR00896

Reichl, J.P.C., Western, A.W., McIntyre, N.R., Chiew, F.H.S., 2009. Optimization of a similarity measure for estimating ungauged streamflow. Water Resour. Res. 45. https://doi.org/10.1029/2008WR007248

Samaniego, L., Bárdossy, A., Kumar, R., 2010. Streamflow prediction in ungauged catchments using copula-based dissimilarity measures. Water Resour. Res. 46. https://doi.org/10.1029/2008WR007695

Samuel, J., Coulibaly, P., Metcalfe, R.A., 2011. Estimation of Continuous Streamflow in Ontario Ungauged Basins: Comparison of Regionalization Methods. J. Hydrol. Eng. 16, 447–459. https://doi.org/10.1061/(asce)he.1943-5584.0000338

Seibert, J., 1997. Estimation of Parameter Uncertainty in the HBV Model. Hydrol. Res. 28, 247-262. https://doi.org/10.2166/nh.1998.15

Seibert, J., Beven, K.J., 2009. Gauging the ungauged basin: How many discharge measurements are needed? Hydrol. Earth Syst. Sci. 13, 883–892. https://doi.org/10.5194/hess-13-883-2009

Sivapalan, M., Takeuchi, K., Franks, S.W., Gupta, V.K., Karambiri, H., Lakshmi, V., Liang, X., McDonnell, J.J., Mendiondo, E.M., O'Connell, P.E., Oki, T., Pomeroy, J.W., Schertzer, D., Uhlenbrook, S., Zehe, E., 2003. IAHS Decade on Predictions in Ungauged Basins (PUB), 2003-2012: Shaping an exciting future for the hydrological sciences. Hydrol. Sci. J. 48, 857–880. https://doi.org/10.1623/hysj.48.6.857.51421

Storn, R., Price, K., 1997. Differential Evolution - A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces. J. Glob. Optim. 11, 341–359. https://doi.org/10.1023/A:1008202821328

Thornton, C.M., Cowie, B.A., Freebairn, D.M., Playford, C.L., 2007. The Brigalow Catchment Study: II. Clearing brigalow (Acacia harpophylla) for cropping or pasture increases runoff. Soil Res. 45, 496. https://doi.org/10.1071/SR07064

Uhlenbrook, S., Seibert, J., Leibundgut, Ch., Rodhe, A., 1999. Prediction uncertainty of conceptual rainfall-runoff models caused by problems in identifying model parameters and structure. Hydrol. Sci. J. 44, 779-797. https://doi.org/10.1080/02626669909492273

Varljen, M.D., Barcelona, M.J., Wehrmann, H.A., 1999. A jackknife approach to examine uncertainty and temporal change in the spatial correlation of a VOC plume. Environ. Monit. Assess. 59, 31–46. https://doi.org/10.1023/A:1006093526872

Vaze, J., Chiew, F.H.S., Perraud, J.M., Viney, N., Post, D., Teng, J., Wang, B., Lerat, J., Goswami, M., 2010. Rainfall-runoff modelling across southeast Australia: Datasets, models and results. Aust. J. Water Resour. 14, 101–116. https://doi.org/10.1080/13241583.2011.11465379

Vogel, R.M., 2005. Regional calibration of watershed models, in: Watershed Models. CRC Press, pp. 47–71. https://doi.org/10.1201/9781420037432.ch3

Wang, R., Kalin, L., 2011. Modelling effects of land use/cover changes under limited data. Ecohydrology 4, 265–276. https://doi.org/10.1002/eco.174

WMO., 2009. Guide to Hydrological Practices, Volume II, Management of Water Resources and

Application of Hydrological Practices, 2009th ed, WMO. WMO, Geneva.

Yang, W., Chen, H., Xu, C.Y., Huo, R., Chen, J., Guo, S., 2020a. Temporal and spatial transferabilities of hydrological models under different climates and underlying surface conditions. J. Hydrol. 591, 125276. https://doi.org/10.1016/j.jhydrol.2020.125276

Yang, X., Magnusson, J., Huang, S., Beldring, S., Xu, C.Y., 2020b. Dependence of regionalization methods on the complexity of hydrological models in multiple climatic regions. J. Hydrol. 582, 124357. https://doi.org/10.1016/j.jhydrol.2019.124357

Yang, X., Magnusson, J., Rizzi, J., Xu, C.Y., 2018. Runoff prediction in ungauged catchments in Norway: Comparison of regionalization approaches. Hydrol. Res. 49, 487–505. https://doi.org/10.2166/nh.2017.071

Yang, X., Magnusson, J., Xu, C.Y., 2019. Transferability of regionalization methods under changing climate. J. Hydrol. 568, 67–81. https://doi.org/10.1016/j.jhydrol.2018.10.030

Yapo, P.O., Gupta, H.V., Sorooshian, S., 1996. Automatic calibration of conceptual rainfall-runoff models: Sensitivity to calibration data. J. Hydrol. 181, 23–48. https://doi.org/10.1016/0022-1694(95)02918-4

Young, A.R., 2006. Stream flow simulation within UK ungauged catchments using a daily rainfall-runoff model. J. Hydrol. 320, 155–172. https://doi.org/10.1016/j.jhydrol.2005.07.017

Zhang, Y., Chiew, F.H.S., 2009. Relative merits of different methods for runoff predictions in ungauged catchments. Water Resour. Res. 45. https://doi.org/10.1029/2008WR007504