

## 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

[Cyswllt i'r cyhoeddiad / Link to publication](#)

*Dyfyniad o'r fersiwn a gyhoeddwyd / Citation for published version (APA):*

Jahanshahi, A., Melsen, L., Patil, S., & Goharian, E. (2021). Comparing spatial and temporal scales of hydrologic model parameter transfer: A guide to four climates of Iran. *Journal of Hydrology*, 603(Part C), Article 127099. <https://doi.org/10.1016/j.jhydrol.2021.127099>

#### Hawliau Cyffredinol / General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

#### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

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

Afshin Jahanshahi\*<sup>1</sup> Lieke A. Melsen<sup>2</sup> Sopan D. Patil<sup>3</sup> Erfan Goharian<sup>4</sup>

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.

## 26    **Abstract**

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

48

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

50

51

## 52 **Introduction**

53 The simulation of streamflow in ungauged catchments remains a challenging task in the hydrologic  
54 sciences (Sivapalan et al., 2003) because the model parameters cannot be calibrated against  
55 streamflow since there are no observations. The process of finding appropriate parameter sets to  
56 simulate streamflow in ungauged sites by learning from model calibration in gauged sites is generally  
57 referred to as “regionalization”. Over the last decade, an increasing number of studies have used  
58 conceptual rainfall-runoff models to test different regionalization approaches (Lee et al., 2005; Merz  
59 and Blöschl, 2004; Perrin et al., 2001; Reichl et al., 2009; Samuel et al., 2011; Vaze et al., 2010;  
60 Vogel, 2005; X. Yang et al., 2020a; Yang et al., 2019). The main issue in regionalization is related to  
61 the operational application of these models outside of calibration periods, where the parameter sets  
62 face their true examination (Dakhlaoui et al., 2017; Patil and Stieglitz, 2015; Refsgaard and Knudsen,  
63 1996; Yang et al., 2020b). Parameter transfer, or regionalization, outside of the calibration period can  
64 be in time (simulating streamflow for periods for which no observations are available), in space  
65 (simulating in ungauged sites) or both (hereafter referred to as “spatiotemporal”). Hrachowitz et al.  
66 (2013), Blöschl et al. (2013), and Parajka et al. (2013) provide a comprehensive overview of the  
67 achievements and discussions in PUB research during the PUB decade initiative (2003-2012)  
68 initiative of the International Association of Hydrological Sciences (IAHS).

69  
70 Temporal transfer of hydrological model parameters is the most common approach in regionalization  
71 studies (Patil and Stieglitz, 2015). An implicit assumption in the temporal transfer is that calibrated  
72 parameters are temporally stable. However, many recent studies have shown that calibrated model  
73 parameters have not been temporally stable (e.g., Brigode et al., 2013; Dakhlaoui et al., 2017; Merz  
74 et al., 2011; Yang et al., 2018) and conditions of calibration period determine their values (Juston et  
75 al., 2009).

76

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; Petheram 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

103 available. Consequently, these studies, especially those with a high number of catchments involved,  
104 have been conducted in regions with a dense and well-organized observation network (the US,  
105 Austria, and France), mainly temperature climates, where the need for regionalization might be  
106 limited. Here, we investigate several methods for regionalization across Iran. Iran is an example of a  
107 country where regionalization might be even more important because of the low gauge density and a  
108 climate regime that is completely different compared to the widely studied in France, Austria, and  
109 the US.

110

111 In Iran, streamflow gauges are generally not in good condition. There are only 1,194 active gauges  
112 and with respect to the total area of the country (1,648,000 km<sup>2</sup>), there is only one active gauge per  
113 1,380 Km<sup>2</sup> (IEM, 2016). The minimum density of streamflow gauges recommended by WMO  
114 (WMO, 2009) is one gauge per 1,875 km<sup>2</sup> and 1,000 km<sup>2</sup> for mountains and interior plains,  
115 respectively. Prediction of streamflow time series in ungauged catchments is a global challenge in  
116 hydrology, and this also applies to Iran - especially in its arid and semi-arid regions. Hence, PUB is  
117 chosen as an essential issue in this study, where we utilize three modes of temporal, spatial, and  
118 spatiotemporal parameter transfer using the available dataset in Iran that covers an extensive range  
119 of climate types.

120

121 To our knowledge, there are only two PUB studies conducted in Iran across the Karkheh River Basin  
122 in the west. Masih et al. (2010) defined hydrological similarity based on four similarity measures:  
123 spatial proximity, drainage area, catchment properties, and Flow Duration Curves (FDC) in 11  
124 ungauged catchments. Their results showed that the physical similarity approach based on similarity  
125 in quantiles of FDC in the HBV model leads to the best performance. In another study, Choubin et  
126 al. (2019) defined the catchment similarity based on morphological, topographic, soil type and land  
127 use, and remote sensing-based characteristics in four catchments. They concluded that physical

128 similarity by applying the semi-distributed SWAT model is an efficient method to estimate  
129 streamflow times series in ungauged catchments.

130

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

136

137 The main questions addressed in this study are:

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

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

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

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

145

## 146 2. Study area, model, and dataset

### 147 2.1. Study area

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

152 There is considerable annual and seasonal variability in rainfall across Iran, with mean annual  
153 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

179 Karkheh River Basin and southwestern Iran (Masih et al., 2011, 2010; Modallakdoust et al., 2008).

180 The equation for this method is as follows:

181 
$$\hat{p}_k = W_D \sum_{i=1}^N \frac{1}{D} w(d)_i p_i + W_Z \sum_{i=1}^N \frac{1}{Z} w(z)_i p_i \quad (1)$$

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

189 
$$w(d) = 1 / d^a \quad \text{for } d > 0 \quad (2)$$

190 
$$w(z) = \begin{cases} 1 / Z_{\min}^b & \text{for } z \leq z_{\min} \\ 1 / z^b & \text{for } z_{\min} < z < z_{\max} \\ 0 & \text{for } z \geq z_{\max} \end{cases} \quad (3)$$

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

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

200 The monthly  $R^2$  (coefficient of determination) ranges from 0.58 to 0.92. Considering the high spatial  
201 variability of precipitation in highlands, the  $R^2$  values are considered satisfactory. A detailed

202 comparison of model efficiency under areal precipitation and gauge (point observations) data is  
203 beyond the scope of this paper. The main parameters used for the interpolation were:  $W_D = 0.8$  and  
204  $W_Z = 0.2$ , radius of influence = 80 km,  $a = 2$ , and  $b = 1$  (Masih et al., 2010). The limiting values for  
205 elevation weighting  $z_{\max}$  and  $z_{\min}$  are selected as 4600 m and 40 m, respectively.

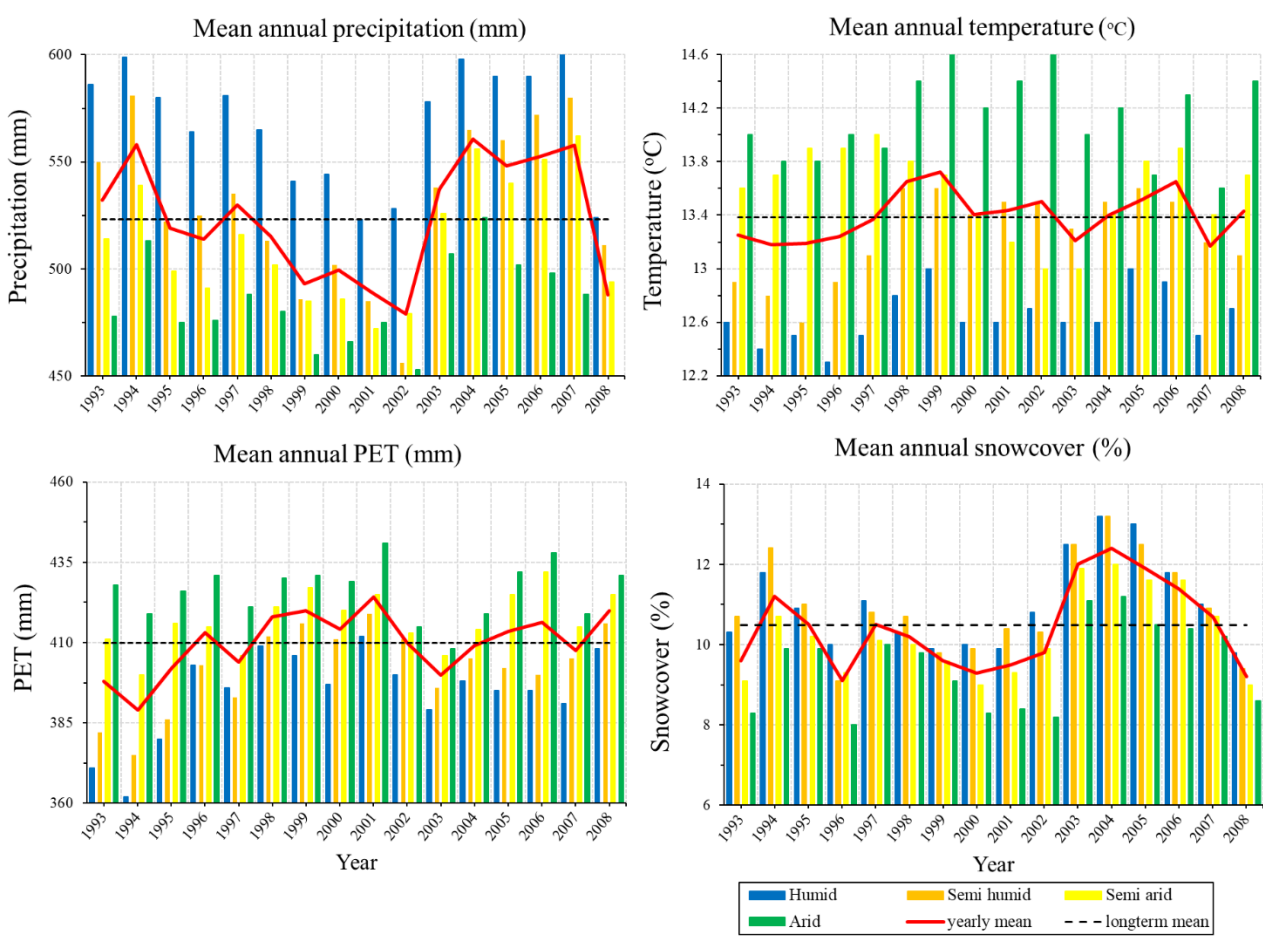
206  
207 (ii) A lapse rate correction ~~is~~ was also tested to account for the elevation effect. Catchment rainfall is  
208 increased by a correction factor that is allowed to vary with mean catchment elevation. This correction  
209 factor was set to 4% in lowland catchments of Iran (below 150 m a.s.l.) and 18 to 25% in the highland  
210 catchments (above 2000 m a.s.l.) and assumed to increase linearly in between.

211  
212 The results showed that the difference between the IDEW and the lapse rate methods, in terms of  
213 catchment averaged rainfall is relatively small (in the order of 3.8%). Hence, the values of catchment  
214 rainfall calculated by the IDEW method are used for this study.

215  
216 There is at least one rain gauge and temperature/evaporation station in each catchment. Daily  
217 temperature time series are generated from the IEM and IMO observations using a regression-based  
218 method by applying elevation as explanatory factor. The reference evapotranspiration is estimated  
219 with the Hargreaves method (Hargreaves et al., 1985) using maximum, minimum, and average  
220 temperature. With the Standard Normal Homogeneity Test (SNHT) (Haimberger, 2007), the time  
221 series of three model inputs was shown to be homogenous, and no breakpoints were observed. The  
222 missing values in the data sets were estimated based on the values from neighboring gauges using the  
223 regression method. Overall, the temperature data of a gauge showed a good correlation with  
224 corresponding data from the neighboring gauges ( $R^2 > 0.89$ ) used for filling the missing records. In  
225 the case of precipitation data, this correlation is  $R^2 > 0.85$ . On average, 7.2% and 10.5% of the  
226 temperature and precipitation data, respectively, had to be filled for all 963 catchments.

227

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



235  
 236 **Fig. 1.** Inter-annual climate variability over the period 1993 to 2008 in the study catchments. The period is  
 237 split into two: 1993-2001 and 2001-2008.

242 **Table 1**

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

Region	Calibration Period	Precipitation (mm)	Temperature (°C)	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

244

## 245 2.4. Catchment dataset

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

255

256 To calibrate and validate the HBV model, daily runoff data from 963 gauged catchments are used  
 257 with areas ranging from 64.7 km<sup>2</sup> to 8432 km<sup>2</sup> and a median of 496 km<sup>2</sup> (Table 2). 97 of these  
 258 catchments range in area between 64.7 and 150 km<sup>2</sup>, 206 catchments between 150 and 350 km<sup>2</sup>, 334  
 259 catchments between 350 and 1000 km<sup>2</sup>, 326 catchments between 1000 and 8432 km<sup>2</sup>. The catchment  
 260 area increases from wet to dry regions so that its median values for humid and semi-arid regions are  
 261 the lowest and highest, respectively (Table 3). The spatial distribution of measuring gauges (in terms  
 262 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  $\geq$  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<sup>2</sup> and 608.2 km<sup>2</sup>, 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).

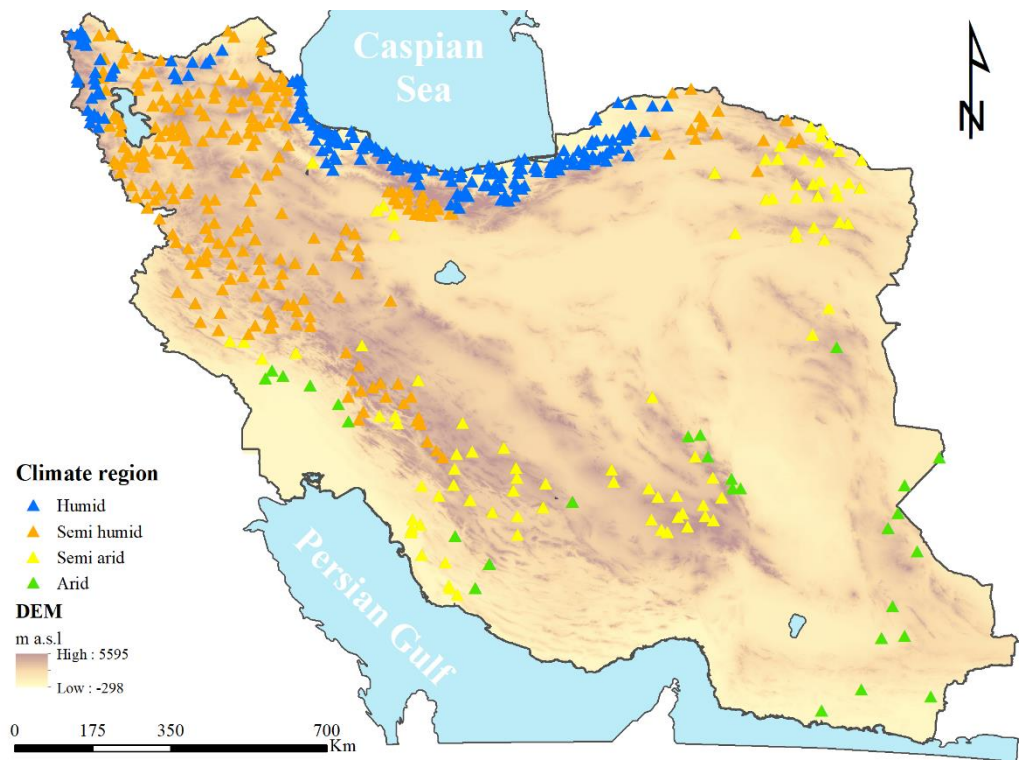
Category	Catchment descriptor	Median	Range	
			Min	Max
Topographic	Mean elevation (m)	1416	-28	5,595
	Mean slope (%)	18.8	0	45.1
Physiographic	Area (km <sup>2</sup> )	574.2	64.7	8,432
Climatic	Aridity Index (PET/P) (-)	0.69	0.1	2.83
	Mean annual precipitation (mm)	724	367	2015
	Mean annual temperature (°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

284 **Table 3**

285 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 <sup>2</sup> )	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

286



287

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

290

291

292

293

### 294 3. Methodology

#### 295 3.1. Model calibration and evaluation

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

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

306

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

310 
$$NSE = \frac{\sum_{i=1}^n (Q_i^{obs} - Q_i^{sim})^2}{\sum_{i=1}^n (Q_i^{obs} - \overline{Q_{obs}})^2} \quad (4)$$

311 where  $Q_i^{sim}$  and  $Q_i^{obs}$  are the daily simulated and observed runoff values at the time  $i$ , respectively.

312  $\overline{Q_{obs}}$  is the mean value of daily observed runoff. One year before each calibration period was used as  
313 the warm-up period to reduce the impact of the uncertain initial conditions on model performance.

314

315

316

### 317 3.2. Parameter transfer strategies and tested modes

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

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

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

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

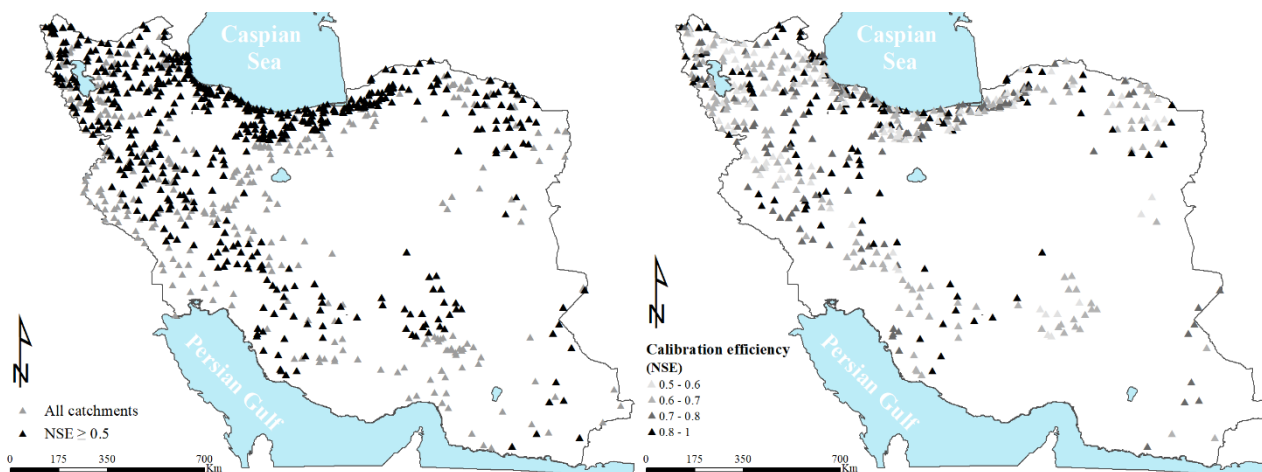
330

## 331 4. Results

### 332 4.1. Model performance over the calibration periods (at-site)

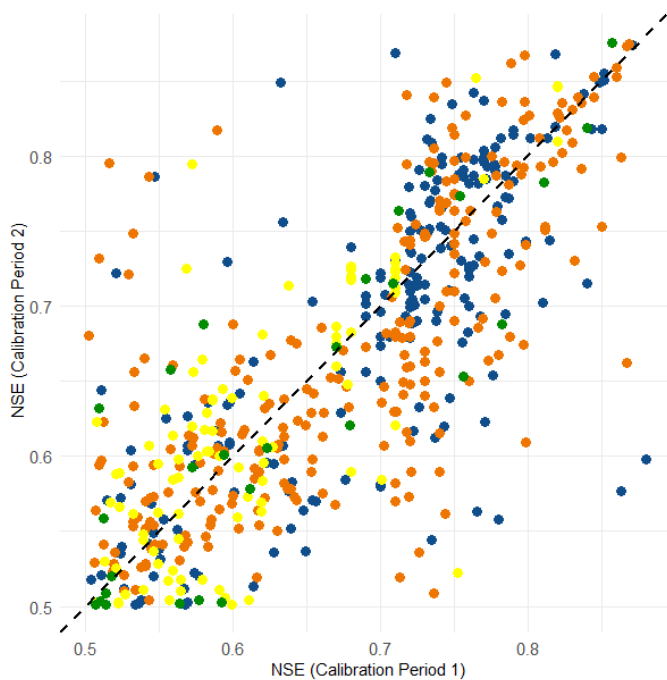
333 We calibrate the model for 963 catchments in two consecutive calibration periods. None of them  
334 showed an NSE lower than 0.29. Subsequently, we eliminate 349 from the initial set of 963  
335 unregulated catchments due to poor model performance for two consecutive periods (calibration NSE  
336  $< 0.5$ ). The discarded catchments (387 catchments with calibration NSE  $< 0.5$ ) are mostly in dry ( $n =$   
337 218) and some in wet ( $n = 169$ ) regions, where karstic aquifers show complex interaction between  
338 surface water and groundwater. Overall, 576 gauged catchments out of the 963 catchments set showed  
339  $NSE \geq 0.5$  for both periods (Fig. 3 left panel). These results, in combination with Fig. 1, show that  
340 the model provides better performance in wetter catchments (humid and semi-humid) than the drier  
341 ones (arid and semi-arid) and are generally in accordance with the findings by Parajka et al. (2005)  
342 in Austria and Oudin et al. (2008) in France.

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



348 **Fig. 3.** Gauge location in Iran. Left panel: Location of all 963 catchments considered within Iran; the 576  
 349 catchments maintained ( $NSE \geq 0.5$ ) for parameter transfer strategy are shown by black triangles, the excluded  
 350 catchments by grey triangles. Right panel: Average calibration efficiency of the two calibration periods over  
 351 the 576 catchments. The triangles have been plotted at the outlet of catchments.  
 352

353  
 354 To compare the calibration results of the two periods, we plotted the NSE values obtained with HBV  
 355 across the two calibration periods in 576 catchments (Fig. 4). The relationship between NSE values  
 356 of these two calibration periods is somewhat weak (Pearson's  $r = 0.55$ ), with fairly widespread data  
 357 points scattered along with both sides of line 1:1. This indicates that model performance can vary  
 358 quite a bit with the calibration period. For all 576 catchments, about 82% of catchments have an NSE  
 359 value of more than 0.6 for two calibration periods. Figure 4 also demonstrates that humid (median  
 360  $NSE = 0.72$ ) and semi-humid (median  $NSE = 0.66$ ) catchments show in general better performance  
 361 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.

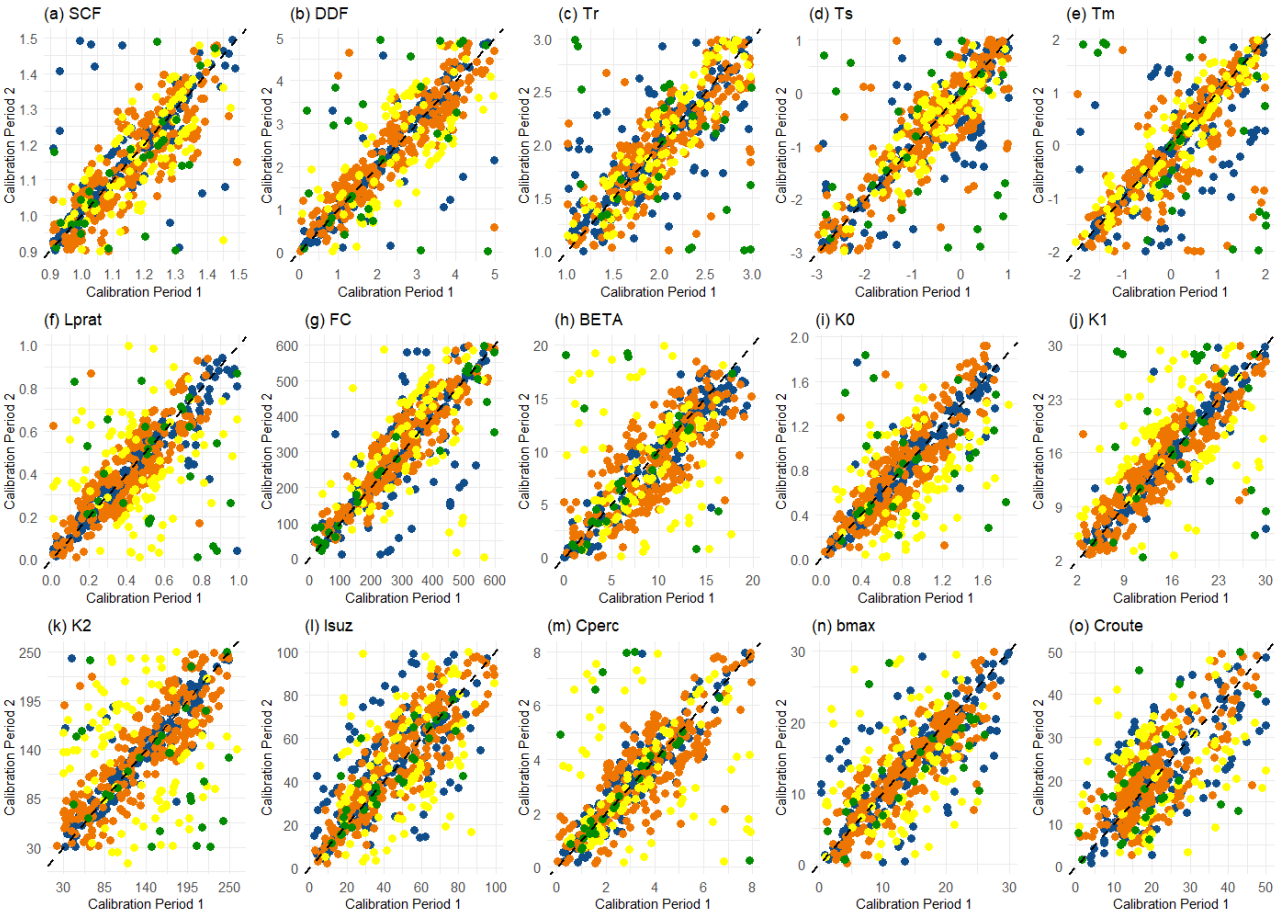
#### 4.2. Parameter uncertainty in regionalization approaches

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 ( $\Delta T$ ) and precipitation ( $\Delta P$ ) 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

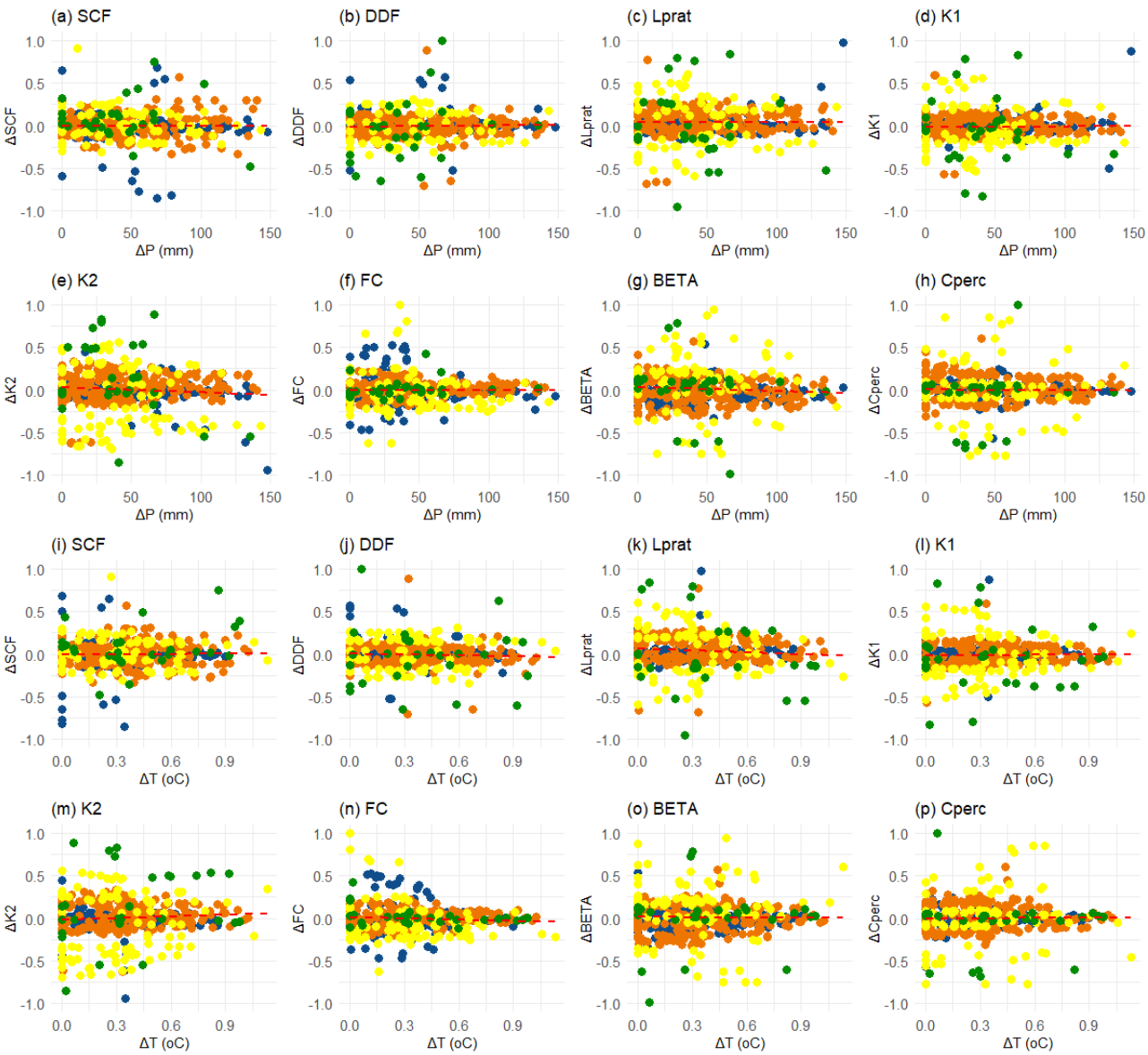
379 stability over time, as indicated by a correlation coefficient of 0.6. The weakest relationship is  
 380 obtained for the storage coefficient for slow response, K2, with a correlation coefficient of 0.29. This  
 381 confirms that not only model performance, as shown in section 4.2, but also parameter values, can  
 382 vary quite a bit across calibration periods. The parameters of the model are less stable for semi-arid  
 383 and arid classes compared to humid and semi-humid classes.



384  
 385 **Fig. 5.** 1:1 plot of all 15 HBV model parameter values for calibration periods. Legend of the climate classes  
 386 according to Fig. 2. Legend of the climate classes according to Fig. 2.

387  
 388 To explore to what extent the variation in parameters during the two calibration periods can be  
 389 attributed to a difference in temperature and precipitation in these two periods, we plotted the change  
 390 in parameter value against  $\Delta P$  and Delta  $\Delta T$  (Fig. 6). It was found that among the 15 calibrated  
 391 parameters (Fig. 5), eight parameters (SCF, DDF, Lprat, K1, K2, FC, BETA, and Cperc) seem to  
 392 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  $\Delta P$  except for SCF, K2, BETA, and Cperc (Fig. 6a to 6h). For  $\Delta T$ , the difference in all parameters shows an increasing trend parallel with an increase of  $\Delta T$  except for DDF, Lprat, FC, and BETA.



**Fig. 6.** 1:1 plot of difference in model parameters ( $\Delta$ Parameter) and (a to h) difference in precipitation ( $\Delta P$ ), (i to p) difference in temperature ( $\Delta T$ ). 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

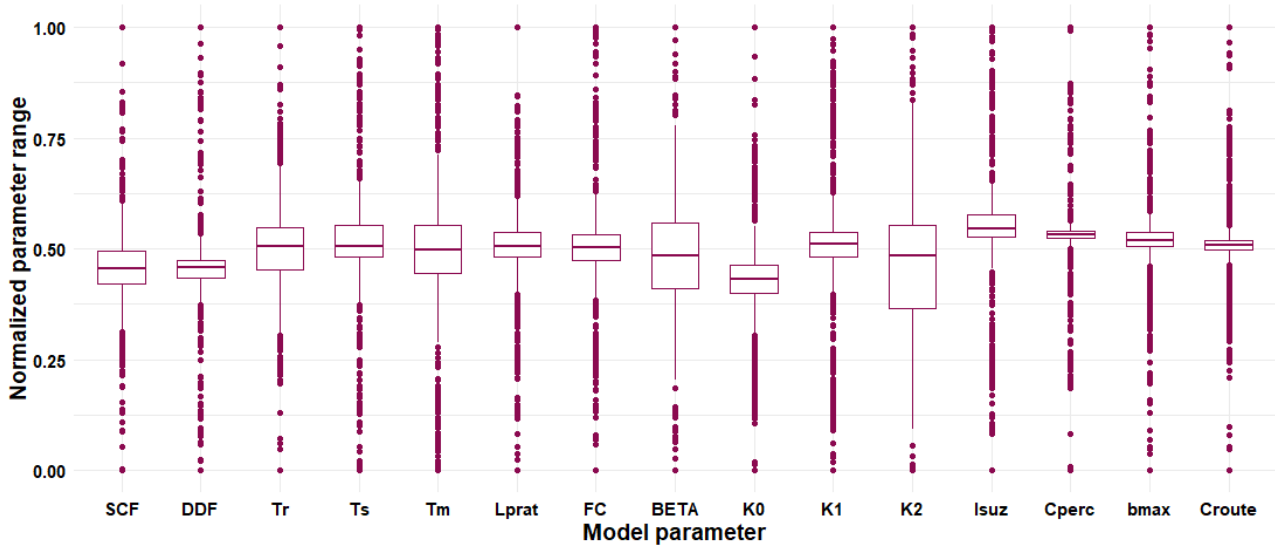
404 evaluate the distance between both values for each catchment. Figure 7 shows the normalized  
 405 parameter range for all 576 catchments. According to this distance distribution, the most and least  
 406 robust parameters are DDF ( $r_2 = 0.61$ ) and Croute ( $r_2 = 0.37$ ) respectively. The  $r_2$  correlation  
 407 coefficients between calibrated and regionalized parameters are presented in Table 4.

408

409 **Table 4**

410 The calibrated parameters of the HBV model, the parameters value range (lower and upper limits), the  
 411 correlation coefficient ( $r_1$ ), median values of parameters for two calibration periods, and the correlation  
 412 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/°C day]	0	5	2.29	0.6	0.61
Tr	Threshold temperature above which precipitation is rain [°C]	1	3	1.98	0.49	0.53
Ts	Threshold temperature below which precipitation is snow [°C]	-3	1.0	-0.68	0.55	0.58
Tm	Threshold temperature above which melt starts [°C]	-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 <sup>2</sup> /mm]	0	50	20.61	0.31	0.37



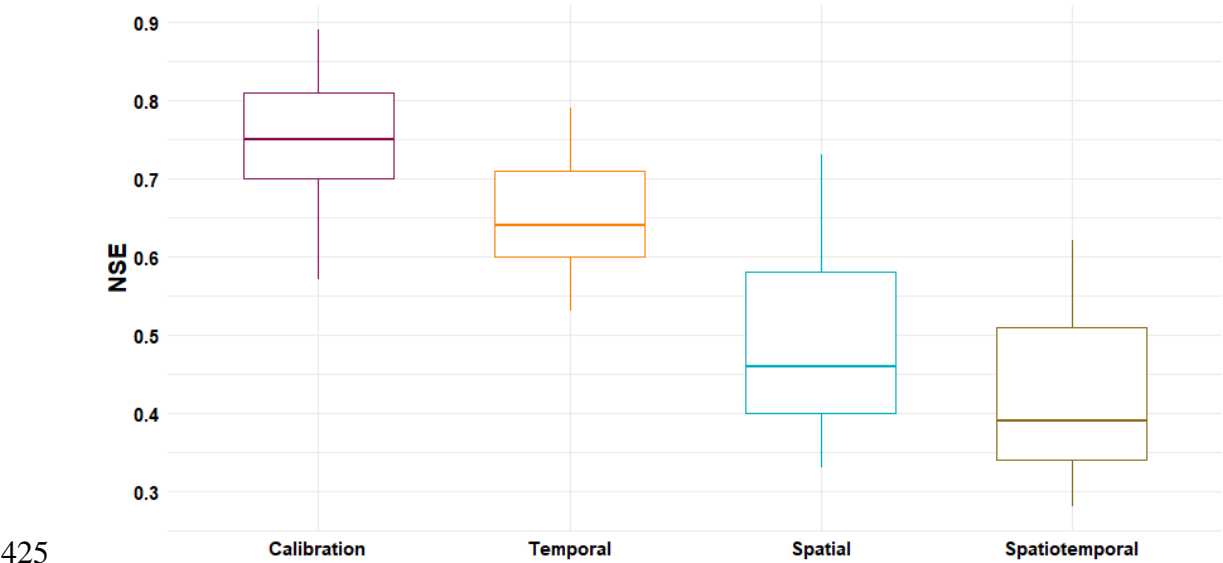
413

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

416

417 4.3. Model performance achieved with the different regionalization approaches

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



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

427

428 4.4. Accounting for temporal and spatial proximities

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

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

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

438

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

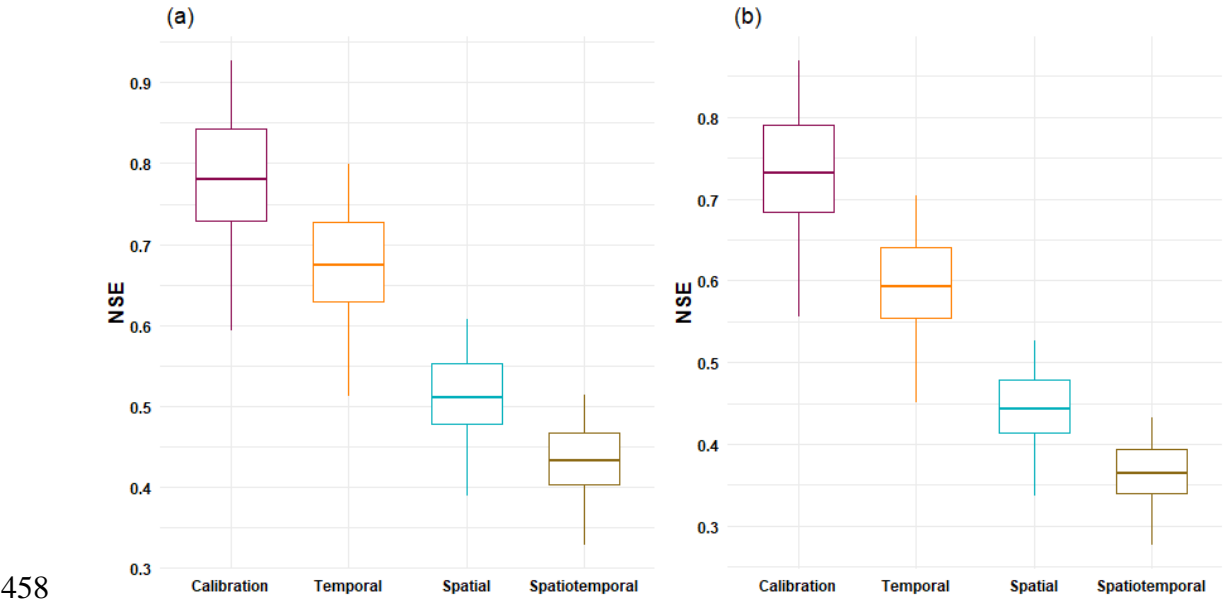
441 *Individual Condition 1:* We remove the catchments that have the nearest neighbor catchment more  
442 than 68 km away (the median nearest neighbor distance is 69.4 km). This reduces the number of  
443 catchments from 576 to 289.

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

450

451 Figures. 9a and 9b show the boxplot of NSE values from four scenarios for individual conditions 1  
452 and 2, respectively. The results for both cases are similar to those results for calibration and parameter  
453 transfer strategies (Fig. 8). For individual condition 1 the highest NSE value is obtained for the  
454 calibration scenario (NSE = 0.78), after that TEM (NSE = 0.67; decline of 13.7%), SPA (NSE = 0.51;  
455 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.



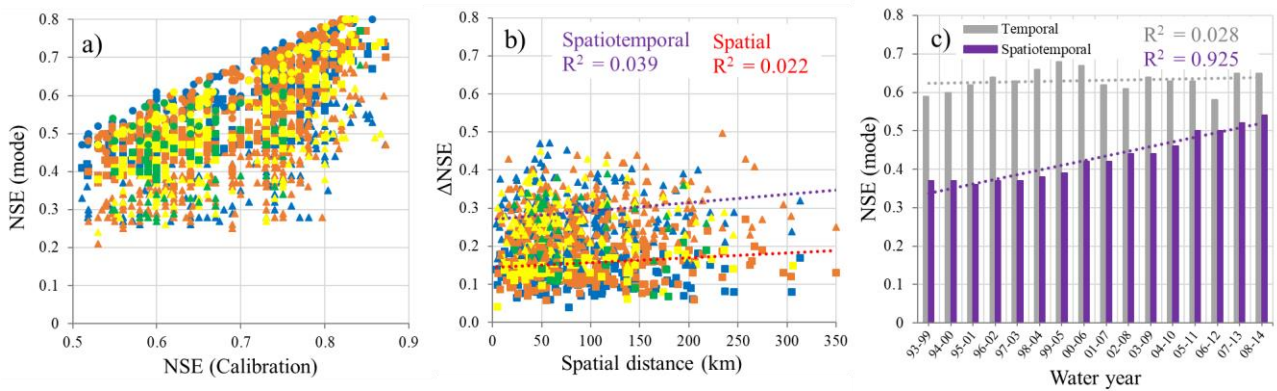
**Fig. 9.** Boxplot of the NSE values for calibration, temporal mode, and two tested strategies. (a) individual condition 1 and (b) individual condition 2.

#### 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 ( $\Delta$ NSE) 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

475 and the spatial distance for both spatial and spatiotemporal. The use of linear regression quantifies  
476 that the effect of distance is somewhat stronger for spatiotemporal ( $R^2 = 0.039$ ) compared to spatial  
477 ( $R^2 = 0.022$ ) (Fig. 10b).  
478  
479 We consider 16 separate temporal lag distances between two calibration periods, so that the  
480 calibration periods 1 changed to 16 WYs from 1993 to 2014 (i.e., 1993-1999, 1994-2000, 1995-2001,  
481 1996-2002, 1997-2003, 1998-2004, 1999-2005, 2000-2006, 2001-2007, 2002-2008, 2003-2009,  
482 2004-2010, 2005-2011, 2006-2012, 2007-2013, and 2008-2014) and calibration period 2 changed to  
483 WY from 2009 to 2014 (instead of 2001-2008). Figure 10c shows the NSE for temporal and  
484 spatiotemporal (y-axis) against NSEs for each WY (x-axis). As seen in Fig. 10c, reducing the  
485 temporal lag between calibration periods reduces the difference between their  
486 temporal/spatiotemporal performance. This is a gradual decline in performance so that the greatest  
487 difference is between WY 1993-1999 and calibration period 2 (2009-2014) for temporal, but the WY  
488 changes to 1995-2001 and 1996-2002 for spatiotemporal. The linear regression for temporal distance  
489 quantifies that the effect of temporal distances is much stronger for spatiotemporal (Fig. 10c) than the  
490 spatiotemporal under spatial distance (Fig. 10b).



491  
492 **Fig. 10.** (a) NSE-calibration versus NSE-mode; (b) NSE-decline ( $\Delta$ NSE) versus spatial distance from donor  
493 catchment for spatial and spatiotemporal and (c) NSE-mode versus the water years between both periods.  
494 Dotted lines indicate the trend lines. Circle, square, and triangle indicate temporal, spatial, and spatiotemporal,  
495 respectively. Legend of the climate classes follows Fig. 2.

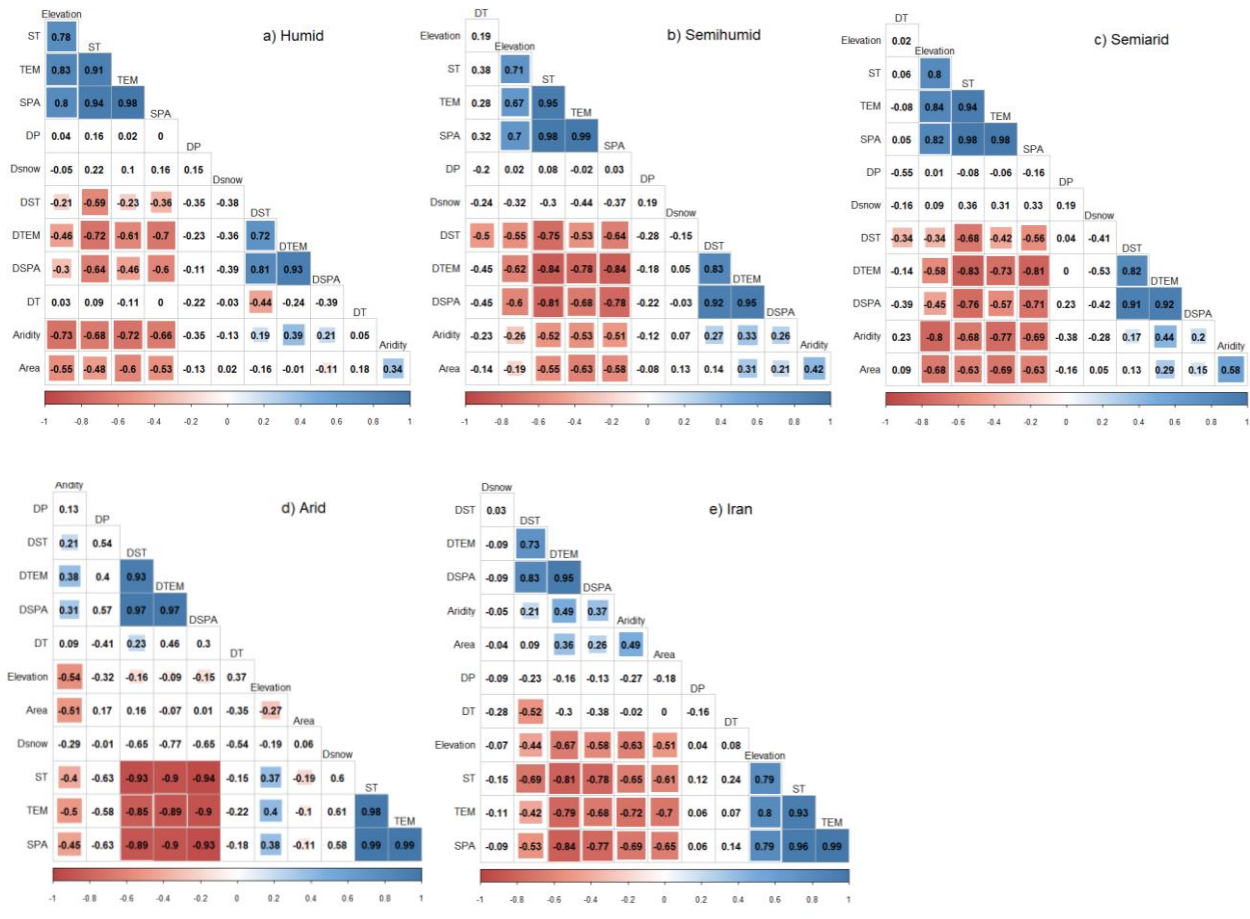
496

#### 497 4.6. Controls on model transferability

498 Figure 11 shows the correlation matrix between catchment characteristics and parameter transfer  
499 strategies. As seen in this Fig., dynamic characteristics have a relatively strong relationship with  
500  $\Delta$ NSE for local and regional scales. The main reason is that precipitation, temperature, and snow  
501 cover changes between the two periods have a remarkable impact on model transferability (Table 3).  
502 Among these dynamic characteristics, aridity has the strongest correlation with  $\Delta$ NSE for both local  
503 and regional scales. This correlation is more significant for the local scale than the regional one.  $\Delta$ P  
504 has a more complicated correlation with  $\Delta$ NSE between two scales. For all three parameter transfer  
505 strategies, its correlation is positive in arid and semi-arid catchments, whereas it is negative in the  
506 humid, semi-humid catchments, and when evaluated at the country level. The correlation between  $\Delta$ T  
507 and  $\Delta$ NSE is negative for humid, semi-humid, semi-arid, and Iran, whereas it is positive for the arid  
508 class.  $\Delta$ snow shows a negative correlation with  $\Delta$ NSE for both regional and local scales. Its  
509 correlation is stronger at the regional scale than the local one.

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

515 In general, correlation coefficients are higher for the temporal transfer strategy than for the spatial  
516 and spatiotemporal strategies, aridity and elevation are the two main controls on model transferability,  
517 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  $\Delta P$ , DT is  $\Delta T$ , Dsnow is  $\Delta snow$ , DTEM, DSPA, and DST are  $\Delta NSE$  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  $\Delta NSE$  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.

549

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.

557

558 Our results showed that Cperc ( $r_1 = 0.36$ ), Croute ( $r_1 = 0.31$ ) and K2 ( $r_1 = 0.29$ ) were known as the  
559 three most sensitive parameters to the changes across the calibration periods. In a more detailed  
560 investigation, the uncertainty of the parameters is investigated by calculating the parameter distance  
561 distribution. The results show that DDF, FC, SCF, and Ts have lower uncertainty, whereas Croute,  
562 BETA, Lprat, and K2 have higher uncertainty. One of the reasons for the high uncertainty of Croute  
563 is that its value tends to be large in lowland and mountainous catchments (about 12%), implying a  
564 more non-linear channel response in these catchments (increasing discharge result in faster response),  
565 but the hydrological reason for these patterns is unclear. This pattern was shown in Merz and Blöschl  
566 (2004) for some of the catchments in northern Austria. Two other parameters (Cperc and K2) are  
567 affected by the runoff generation conditions of the catchments, not the input data conditions during  
568 two calibration periods. The most stable parameters in terms of temporal are DDF ( $r_1 = 0.6$ ), FC ( $r_1$   
569  $= 0.58$ ) and SCF ( $r_1 = 0.56$ ) which represent: (1) degree day factor, (2) field capacity and (3) snow  
570 correction factor which are unlikely to be affected by severe temporal changes during calibration  
571 periods. These three parameters are affected by snow and soil conditions of calibration periods and  
572 studied catchments. Generally, the parameters related to snow and runoff routines are the lowest and  
573 highest uncertain parameters compared to the other model routines.

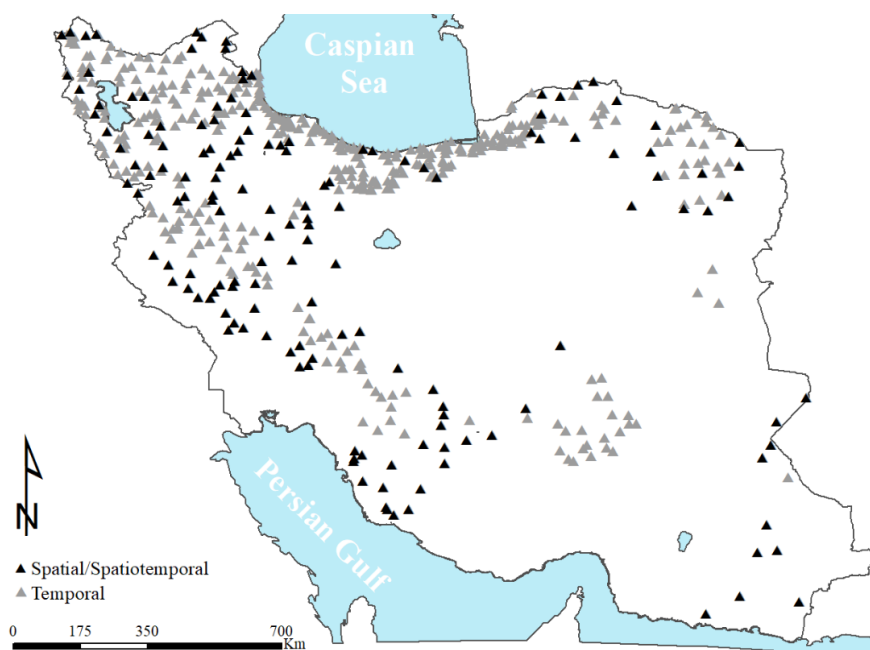
574

575 Nevertheless, the parameters, which are stables in terms of temporal variability, are different for  
576 different types of the rainfall-runoff model used. Patil and Stieglitz (2015) calibrated EXP-HYDRO  
577 in 294 catchments in the U.S. for two calibration periods. They concluded that the TEM variability is  
578 different between all parameters. In their study, the parameter  $f$ , showed the lowest uncertainty  
579 (highest correlation) between calibration periods (lowest uncertainty). Merz and Blöschl (2004)  
580 calibrated the HBV model in 308 catchments of Austria for two calibration periods and founded that  
581 the range of  $R^2$  between two sets of calibrated parameters ranged between 0.09 and 0.64 (only 5 out  
582 of 11 parameters have  $R^2 \geq 0.5$ ). Oudin et al. (2008) compared TOPMO (6 parameters) and GR4J (4  
583 parameters) models for 913 catchments in France. Their results showed the superiority of the GR4J

584 model compared to TOPMO in terms of a higher correlation between the parameters across  
585 calibration periods. Although, it is difficult to judge how the model parameters values (for all  
586 hydrological rainfall-runoff models used in the literature on the model parameters transfer) will  
587 change in response to wet and dry periods and land-use changes within catchments. This insight  
588 requires more extensive and comprehensive research (Eckhardt et al., 2003; Patil and Stieglitz, 2015;  
589 Wang and Kalin, 2011). Thus, even when considering variable parameters (Section 4.2.1), model  
590 transferability will still be prone to error. The error became larger when study catchments behaved  
591 differently under extreme events. Other possible effects of climate variability in the long-term are  
592 changes in vegetation and the water table that need to be assessed in data-sparse catchments of Iran.

593  
594 Our results from the parameter transfer strategies (Fig. 8) show the overall superiority of the temporal  
595 (TEM) transfer mode over spatial (SPA) and spatiotemporal (SPA\_TEM) transfer strategies.  
596 Individual comparison of them across all 576 catchments clarified that the TEM mode has the best  
597 performance at 414 catchments (and worst at 88 catchments), the SPA strategy is best in 118  
598 catchments (and worst at 219 catchments), and the SPA\_TEM strategy is best in 44 catchments (and  
599 worst at 269 catchments). Figure 12 shows the location of catchments, where either the SPA or  
600 SPA\_TEM is the best case. No specific and regular geographic pattern is deduced from the spatial  
601 distribution of tested catchments in terms of the superiority of SPA and SPA\_TEM strategies over  
602 the TEM case. Table 5 shows the comparison of two groups of transfer strategies between studied  
603 catchments (Group 1: TEM mode performing best; Group 2: SPA or SPA\_TEM strategy performing  
604 best) in terms of three hydroclimatic indices: aridity index ( $PET/P$ ), annual runoff ratio ( $Q/P$ ) and  
605 mean annual rainfall ( $P$ ). Even though the median values of these three indices show that the  
606 catchments in group 1 are wetter (lower and higher values of  $PET/P$  and  $P$ , respectively) and less  
607 flashy (lower  $Q/P$ ). Nevertheless, Fig. 12 shows in some parts of the study area with a low density of  
608 catchments and larger distance between neighboring catchments (e.g., in the central, southwestern  
609 and southeastern parts of Iran), the TEM mode is not superior to SPA and SPA\_TEM strategies. This

610 relatively irregular geographical pattern was also demonstrated in the U.S. by Patil and Stieglitz  
 611 (2015), who found that in regions with low catchment density, the temporal mode of parameter  
 612 transfer does not always outperform the other two strategies.



613  
 614 **Fig. 12.** Location of catchments where the temporal mode performs best (grey triangles) and catchments where  
 615 either spatial or spatiotemporal perform best (black triangles). The triangles have been plotted at the outlet of  
 616 catchments.

617

618 **Table 5**

619 Median values of three hydro-climatic descriptors for two catchment groups (shown by Fig. 12) (n = 576).  
 620 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)

621

622 The evaluation of characteristics (aridity, differences in precipitation, temperature and snow cover of  
 623 calibration and validation periods, catchment area and elevation) on model transferability showed  
 624 that the slight climatic variability between the two calibration periods has a remarkable effect on  
 625 model transferability. Our result shows that aridity and catchment elevation are the two major controls  
 626 on model transferability at two regional (climate classes) and local (the whole country) scales. These

627 effects are most robust at a regional scale. This finding shows that negligible climatic variability  
628 affects the model transferability more at a smaller scale than a larger scale. The more noticeable effect  
629 of static characteristics (catchment area and elevation) on model transferability indicates that these  
630 two physical characteristics are useful descriptors for model transferability.

631 Under the two individual conditions, model transferability shows that the TEM mode retained its  
632 superiority over the SPA and SPA\_TEM strategies. For *Individual Condition 1*, when only those  
633 catchments with the nearest neighbor < 68 km away are maintained, the SPA and SPA\_TEM  
634 strategies showed performance improvement compared to the base scenario, so that their differences  
635 with calibration is about 3.59% lower than the base scenario (in terms of NSE value). This is an  
636 expected result because reducing the spatial distance between donor and target catchments will most  
637 likely reduce the spatial variability of hydrological behavior and improve performance (Oudin et al.,  
638 2008; Patil and Stieglitz, 2015). This finding is confirmed by testing the relationship between  $\Delta$ NSE  
639 and spatial distance (Fig. 10b) so that increasing spatial distance between donor and target catchments  
640 increases  $\Delta$ NSE for SPA and SPA\_TEM (see section 4.5). For *Individual Condition 2* (when a 10-  
641 year temporal gap is added between two calibration periods), the NSE difference between calibration  
642 and the TEM mode is 4.51% higher than the base scenario and is virtually unchanged between  
643 calibration and two other strategies.

644

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

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

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

657  
658 **5. Conclusion**

659 This study, which is so far the most comprehensive PUB study in Iran, investigated three strategies  
660 for transferring model parameters, including temporal, spatial, and spatiotemporal, during two  
661 calibration periods using a conceptual rainfall-runoff model called HBV at 576 unregulated  
662 catchments across Iran. Our results showed that temporal mode has the best performance, with the  
663 lowest decline in performance than calibration (median decline of 14.66%), while these declines for  
664 spatial and spatiotemporal were 38.6% and 48%, respectively. Thus, we conclude that the stability of  
665 the parameters of the HBV model in the temporal mode is higher than for spatial strategy. Hence,  
666 temporal comes out best, making sense because the model has “seen” the catchment. This finding is  
667 in accordance with previous studies by Zhang and Chiew (2009) in Australia, Parajka et al. (2005) in  
668 Austria, and Patil and Stieglitz (2015) in the U.S. We also showed that the superiority of temporal  
669 mode is maintained under three scenarios ((1) a decrease in the spatial (geographical) distance  
670 between donor and target catchments; and (2) an increase in the temporal lag (10 years) between  
671 calibration and validation periods and (3) a gradual increase in the temporal lag between calibration  
672 and validation periods). This finding is consistent with a previous study by Patil and Stieglitz (2015),  
673 but with relatively poorer performance at both temporal and spatial transfer strategies. We also  
674 concluded that an increase in temporal lag between calibration and validations leads to a reduction in  
675 the model performance gap between the temporal mode and spatial and spatiotemporal strategies.  
676 This finding suggests that spatiotemporal parameter transfer can be a fairly reliable option for PUB  
677 studies and climate change-related studies, at least in wetter catchments. Our results are obtained  
678 from two consecutive calibration periods (by averaging model parameters). However, more research

679 is needed to: (i) examine longer calibration periods, (ii) obtain a stronger relationship between  
680 temporal mode and spatial strategy, (iii) explore the effects of increasing time lag (>10 years) between  
681 calibration and validation periods on results to achieve more confident about the variability of the  
682 hydrological model parameters. Exploring dynamic (aridity and differences in precipitation,  
683 temperature and snow cover of calibration and validation periods) and static (catchment area and  
684 mean elevation) controls on model transferability showed that aridity and catchment elevation are  
685 two major controls on model transferability at two regional (climate classes) and local (the whole  
686 country) scales.

687

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

692

## 693 **Acknowledgments**

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

698

## 699 **References**

- 700 Ardia, D., Boudt, K., Carl, P., Mullen, K.M., Peterson, B.G., 2011. Differential evolution with  
701 deoptim. R J. 3, 27–34. <https://doi.org/10.32614/rj-2011-005>
- 702 Arsenault, R., Brissette, F.P., 2014. Continuous streamflow prediction in ungauged basins: The  
703 effects of equifinality and parameter set selection on uncertainty in regionalization approaches.  
704 Water Resour. Res. 50, 6135–6153. <https://doi.org/10.1002/2013WR014898>

705 Bao, Z., Zhang, J., Liu, J., Fu, G., Wang, G., He, R., Yan, X., Jin, J., Liu, H., 2012. Comparison of  
706 regionalization approaches based on regression and similarity for predictions in ungauged  
707 catchments under multiple hydro-climatic conditions. *J. Hydrol.* 466–467, 37–46.  
708 <https://doi.org/10.1016/j.jhydrol.2012.07.048>

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

711 Bergström, S., 1976. Development and application of a conceptual runoff model for Scandinavian  
712 catchments. Series A, Lund Institute of Technology,. University of Lund, Lund, Sweden. No.  
713 52.

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

717 Brigode, P., Oudin, L., Perrin, C., 2013. Hydrological model parameter instability: A source of  
718 additional uncertainty in estimating the hydrological impacts of climate change? *J. Hydrol.* 476,  
719 410–425. <https://doi.org/10.1016/j.jhydrol.2012.11.012>

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

725 Choubin, B., Solaimani, K., Rezanezhad, F., Habibnejad Roshan, M., Malekian, A., Shamshirband,  
726 S., 2019. Streamflow regionalization using a similarity approach in ungauged basins:  
727 Application of the geo-environmental signatures in the Karkheh River Basin, Iran. *Catena* 182,  
728 104128. <https://doi.org/10.1016/j.catena.2019.104128>

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

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

733 Dakhlaoui, H., Ruelland, D., Tramblay, Y., Bargaoui, Z., 2017. Evaluating the robustness of  
 734 conceptual rainfall-runoff models under climate variability in northern Tunisia. J. Hydrol. 550,  
 735 201–217. <https://doi.org/10.1016/j.jhydrol.2017.04.032>

736 De Martonne E., 1926. Une nouvelle fonction climatologique: L'indice d'aridité. La Meteorologie,  
 737 2:449–458

738 Eckhardt, K., Breuer, L., Frede, H.G., 2003. Parameter uncertainty and the significance of simulated  
 739 land use change effects. J. Hydrol. 273, 164–176. [https://doi.org/10.1016/S0022-](https://doi.org/10.1016/S0022-1694(02)00395-5)  
 740 1694(02)00395-5

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

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

746 Hargreaves, G.L., Hargreaves, G.H., Riley, J.P., 1985. Irrigation Water Requirements for Senegal  
 747 River Basin. J. Irrig. Drain. Eng. 111, 265–275. [https://doi.org/10.1061/\(asce\)0733-](https://doi.org/10.1061/(asce)0733-9437(1985)111:3(265))  
 748 9437(1985)111:3(265)

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

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

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

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

759 Juston, J., Seibert, J., Johansson, P.-O., 2009. Temporal sampling strategies and uncertainty in  
 760 calibrating a conceptual hydrological model for a small boreal catchment. *Hydrol. Process.* 23,  
 761 3093–3109. <https://doi.org/10.1002/hyp.7421>

762 Kay, A.L., Jones, D.A., Crooks, S.M., Kjeldsen, T.R., Fung, C.F., 2007. An investigation of site-  
 763 similarity approaches to generalisation of a rainfall-runoff model. *Hydrol. Earth Syst. Sci.* 11,  
 764 500–515. <https://doi.org/10.5194/hess-11-500-2007>

765 Kennard, M.J., Pusey, B.J., Olden, J.D., MacKay, S.J., Stein, J.L., Marsh, N., 2010. Classification of  
 766 natural flow regimes in Australia to support environmental flow management. *Freshw. Biol.* 55,  
 767 171–193. <https://doi.org/10.1111/j.1365-2427.2009.02307.x>

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

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

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

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

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

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

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

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

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

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

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

796 Nash, J.E., Sutcliffe, J. V., 1970. River flow forecasting through conceptual models part I - A  
797 discussion of principles. *J. Hydrol.* 10, 282–290. [https://doi.org/10.1016/0022-1694\(70\)90255-](https://doi.org/10.1016/0022-1694(70)90255-6)  
798 6

799 Osuch, M., Wawrzyniak, T., Nawrot, A., 2019. Diagnosis of the hydrology of a small Arctic  
800 permafrost catchment using HBV conceptual rainfall-runoff model. *Hydrol. Res.* 50, 459–478.  
801 <https://doi.org/10.2166/nh.2019.031>

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

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

808 Parajka, J., Merz, R., Blöschl, G., 2005. A comparison of regionalisation methods for catchment

809 model parameters. *Hydrol. Earth Syst. Sci.* 9, 157–171. <https://doi.org/10.5194/hess-9-157-2005>

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

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

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

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

820 Patil, S.D., Wigington, P.J., Leibowitz, S.G., Comeleo, R.L., 2014a. Use of hydrologic landscape  
821 classification to diagnose streamflow predictability in Oregon. *J. Am. Water Resour. Assoc.* 50,  
822 762–776. <https://doi.org/10.1111/jawr.12143>

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

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

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

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

835 97a6-83d715506c56 (accessed 3.31.21).

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

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

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

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

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

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

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

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

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

860 Seibert, J., Beven, K.J., 2009. Gauging the ungauged basin: How many discharge measurements are

861 needed? Hydrol. Earth Syst. Sci. 13, 883–892. <https://doi.org/10.5194/hess-13-883-2009>  
 862 Sivapalan, M., Takeuchi, K., Franks, S.W., Gupta, V.K., Karambiri, H., Lakshmi, V., Liang, X.,  
 863 McDonnell, J.J., MENDIONDO, E.M., O’Connell, P.E., Oki, T., Pomeroy, J.W., Schertzer, D.,  
 864 Uhlenbrook, S., Zehe, E., 2003. IAHS Decade on Predictions in Ungauged Basins (PUB), 2003-  
 865 2012: Shaping an exciting future for the hydrological sciences. Hydrol. Sci. J. 48, 857–880.  
 866 <https://doi.org/10.1623/hysj.48.6.857.51421>  
 867 Storn, R., Price, K., 1997. Differential Evolution - A Simple and Efficient Heuristic for Global  
 868 Optimization over Continuous Spaces. J. Glob. Optim. 11, 341–359.  
 869 <https://doi.org/10.1023/A:1008202821328>  
 870 Thornton, C.M., Cowie, B.A., Freebairn, D.M., Playford, C.L., 2007. The Brigalow Catchment  
 871 Study: II. Clearing brigalow (*Acacia harpophylla*) for cropping or pasture increases runoff. Soil  
 872 Res. 45, 496. <https://doi.org/10.1071/SR07064>  
 873 Uhlenbrook, S., Seibert, J., Leibundgut, Ch., Rodhe, A., 1999. Prediction uncertainty of conceptual  
 874 rainfall-runoff models caused by problems in identifying model parameters and structure. Hydrol.  
 875 Sci. J. 44, 779-797. <https://doi.org/10.1080/02626669909492273>  
 876 Varljen, M.D., Barcelona, M.J., Wehrmann, H.A., 1999. A jackknife approach to examine uncertainty  
 877 and temporal change in the spatial correlation of a VOC plume. Environ. Monit. Assess. 59, 31–  
 878 46. <https://doi.org/10.1023/A:1006093526872>  
 879 Vaze, J., Chiew, F.H.S., Perraud, J.M., Viney, N., Post, D., Teng, J., Wang, B., Lerat, J., Goswami,  
 880 M., 2010. Rainfall-runoff modelling across southeast Australia: Datasets, models and results.  
 881 Aust. J. Water Resour. 14, 101–116. <https://doi.org/10.1080/13241583.2011.11465379>  
 882 Vogel, R.M., 2005. Regional calibration of watershed models, in: Watershed Models. CRC Press, pp.  
 883 47–71. <https://doi.org/10.1201/9781420037432.ch3>  
 884 Wang, R., Kalin, L., 2011. Modelling effects of land use/cover changes under limited data.  
 885 Ecohydrology 4, 265–276. <https://doi.org/10.1002/eco.174>  
 886 WMO., 2009. Guide to Hydrological Practices, Volume II, Management of Water Resources and

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

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

891 Yang, X., Magnusson, J., Huang, S., Beldring, S., Xu, C.Y., 2020b. Dependence of regionalization  
892 methods on the complexity of hydrological models in multiple climatic regions. *J. Hydrol.* 582,  
893 124357. <https://doi.org/10.1016/j.jhydrol.2019.124357>

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

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

899 Yapo, P.O., Gupta, H.V., Sorooshian, S., 1996. Automatic calibration of conceptual rainfall-runoff  
900 models: Sensitivity to calibration data. *J. Hydrol.* 181, 23–48. [https://doi.org/10.1016/0022-](https://doi.org/10.1016/0022-1694(95)02918-4)  
901 [1694\(95\)02918-4](https://doi.org/10.1016/0022-1694(95)02918-4)

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

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