Identifying most relevant controls on catchment hydrological similarity using model transferability - A comprehensive study in Iran

Jahanshahi, Afshin; Patil, Sopan; Goharian, Erfan

Journal of Hydrology

DOI:
https://doi.org/10.1016/j.jhydrol.2022.128193

Published: 01/09/2022

Identifying most relevant controls on catchment hydrological similarity using model transferability - A comprehensive study in Iran

Afshin Jahanshahi*[1] Sopan D. Patil[2] Erfan Goharian[3]

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

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

Abstract

A key goal of hydrological science is to understand the climatic controls on catchment hydrological response behavior. Progress toward this goal would result in improved model transferability from gauged to gauged catchments. In this sense, the effectiveness of the model's transferability is contingent on the proper selection of donor and target catchment pairs. Thus, using a rainfall-runoff model, we evaluate two distinct types of hydrological similarity in this study: (i) the apparent similarity measured by similarity distance based on observable catchments descriptors (CDs) and a Euclidean distance based on physical similarity (PS) method and (ii) behavioral similarity, which is determined by highest-performance of transferred model parameters between gauged donor catchment and ungauged target catchment (best-donor case (BD)). It is believed that catchments that apparently to be similar in terms of CDs, have a similar hydrological behavior. We wish to see if that assumption is valid in this paper. Spatial proximity (SP) is also implemented to see if it might be used as an alternative for PS where there is no apparent physical similarity between catchments. To test the study's assumptions, the HBV conceptual rainfall-runoff model is used in 576 catchments across four climate regions in Iran. The results indicate that: (1) as expected, the best-donor (BD) case performs the best, and the more than 75% of our physically similar catchments have a hydrological similarity (the overlap was $\geq$ 70%), (2) the superiority of PS over SP demonstrates that the CDs exert a great influence on transferability within each climate region than geographical distance. However, we demonstrated that the SP is superior, when spatial distance between donor and target catchments is reduced (nearest neighbor $\leq$ 20 km), (3) consistent with CDs, when utilizing SP method, geographical distance has a varying effect on model transferability within wetter and drier regions, such that SP performs better in wetter regions than it does in dry interior regions, (4) throughout Iran, the dominant controls on model transferability differ by region. Thus, the climatic (aridity index or PET/P), topographic (mean elevation), and physiographic (catchment area) properties exert a greater influence on parameter transfer to ungauged catchments than do other CDs, and (5) the runoff ratio

(streamflow signature) confirmed the superiority of the wetter regions over the drier regions in terms of control on the parameters transfer.

Keyword: Apparent similarity, Behavioral similarity, Geographical distance, Physical similarity, Spatial proximity.

Introduction

Accurate estimation of hydrological model parameters requires information about continuous streamflow time series for river gauges, but this data is incomplete and unavailable in many river basins of the world (Blöschl et al., 2013; Oudin et al., 2010). Regionalization is therefore an important issue in hydrological science (Sivapalan, 2003). We define regionalization in this study as all methods that allow hydrological information to be transferred from gauged to ungauged catchments.

Obtaining model parameters from supposedly similar catchments is an appealing approach to estimate model parameters in ungauged catchments. This is the rationale for the physical similarity approach, which seeks to detect hydrological similarities based on the physical characteristics of catchments. (Burn and Boorman, 1993). Apart from the arduous task of understanding which catchment descriptors (CDs) influence hydrological behavior, there is also the issue of defining hydrological similarity. Indeed, the hydrological similarity is frequently determined by examining the signatures of catchment functional responses (e.g., runoff yield), which are influenced to some extent by climate variables (Oudin et al., 2010).

Currently, similarity-based approaches established in regionalization studies aim to transfer a model parameter set calibrated on a gauged donor catchment to the target ungauged catchment if the donor catchment is physically similar to the target catchment (McIntyre et al., 2005). Two significant assumptions are made implicitly in this procedure: (1) because the calibrated model parameter sets acquired from two different catchments are similar, it is expected that their behavior in response to the transformation of rainfall-runoff is similar, and (2) it is assumed that a catchment's physical similarity (as determined by multiple CDs) reflects a hydrological similarity between the two.

From a hydrological standpoint, assumption 1 is inescapable. It does, however, have limitations, the first of which is the potential of compensations between parameters (Kokkonen et al., 2003). Over-parameterized models are more susceptible to this issue than other types of models. Secondly, while calibrated parameters represent catchment behavior, they can also reveal biases in the data used to calculate them (Andréassian et al., 2004, 2001; Oudin et al., 2006).

Every regionalization study has Assumption 2. In summary, it assumes that it should be feasible to find specific CDs that account for the hydrological catchment behavior. Given the potential value of various CDs in defining catchments' hydrological response, it is necessary to ascertain which CDs are most appropriate (Wagener et al., 2007), particularly for regionalization purposes (Blöschl, 2005).

This study aims to identify which gauged catchment(s) are hydrologically most similar to the ungauged target catchment in order to find out how much readily available CDs explain catchment hydrological behavior and hence which are the best donors of model parameters. Thus, employing the relevant CDs to establish an appropriate similarity between gauged and ungauged catchments (apparent similarity), a better presentation of the hydrologic similarity (HS) can be guaranteed. This assumption is similar to those considered by Falkenmark and Chapman (1989), Sivapalan (2009), and Oudin et al. (2010).

By analyzing different catchments located in different climate regions, a framework for identifying the most relevant controls can be developed. Several studies have been conducted in the literatures that have considered this hypothesis and yielded interesting results. For example, in a comprehensive study, Parajka et al. (2013) reviewed many studies on a wide range of regionalization approaches and concluded that it is easier to simulate streamflow in larger catchments than in smaller ones, and in humid catchments as to compared to dry ones. In 913 French catchments, Oudin et al. (2008) conclude that CDs (climatic, physiographic, and topographic) may vary by region and obtaining a unique CD or consistent set of CDs to define HS between donor and target catchment may not be possible. They determined that in another study that there is a high overlap between apparent (physical) and behavioral (hydrological) similarity for 60% of French and UK catchments (Oudin et al., 2010). In 83 catchments across the United States, Singh et al. (2014) concluded that the optimal CDs for defining hydrologic similarity differ by region, with physical/climatic characteristics identified as the most relevant CDs for transferring hydrologic model parameters. They also concluded that regional analysis of regionalization approaches can aid in identifying the most relevant controls on parameter transfer for each hydro climate region.

Unlike earlier PUB research (Merz and Blöschl, 2004; Oudin et al., 2008; Parajka et al., 2005; Patil and Stieglitz, 2015; Zhang and Chiew, 2009) we identified a case of highest-performance of transferred model parameters between gauged donor catchment and ungauged target catchments, regardless of spatial distance, climatic type and/or physical properties (best-donor case (BD)). The term "behavioral similarity" is referred to this. Despite two earlier comprehensive investigations of model transferability in Iran (Jahanshahi et al., 2022, 2021), this can serve as a baseline or reference score for interpreting model transferability when apparent similarity exists.

We conduct a comprehensive analysis on two scales of (a) climate regions (regional) and (b) throughout Iran (local). The Similarity Index (SI) is constructed based on a variety of combinations of (a) dynamic (climate) and (b) static (physiographic and land use) CDs in Physical Similarity (PS) regionalization method. Spatial proximity (SP) is also implemented to see if it might be used as an alternative for PS where there is no apparent physical similarity between catchments. The geographical distance (GD) between gauged and ungauged catchments is employed as a similarity metric for parameter set transfer in this method (e.g., Yang et al., 2020; Jahanshahi et al. 2021).

Thus, the degree of relevance of each CDs to regionalization is determined in this study, as the first and most comprehensive study on identifying the main controls on HS at the national scale of Iran.

This study aims to answer three main questions:

(a) Do apparent physical similar catchments have a similar hydrological behavior?

(b) Which metric (or metrics) cause the high performance of HS (parameter transfer approaches): physical, climatic, geographical, or a combination of the three?

(c) Is dominant control(s) on the transferring of hydrologic model parameters the same across different types of climate regions?

Our innovation in this study are: (a) we moved beyond a variety of single and groups of dynamic (climate) and static (physiographic, land use, and geographical distance) catchment descriptors to establish the optimal case(s) of hydrologic similarity in terms of parameter transfer performance and then compared it to the highest-performance case of parameter transfer. This is beneficial because it connects complex catchment responses to relevant catchment descriptors, and (b) we also used information about streamflow signature (runoff ratio) to interpret parameter transfer results.

2. Study area, model, and dataset

2.1. Study area

Iran is our study area. According to the De Martonne classification system (De Martonne, 1926; Jahanshahi et al., 2021), Iran is divided into four major climate regions,. The country's climate varies greatly, ranging from a humid and semi-humid maritime climate along the Caspian Sea coast to arid and semi-arid climates in the interior.

Annual and seasonal rainfall in Iran is highly variable, with mean annual precipitation (MAP) ranging from 360 mm in the middle regions to more than 2000 mm in the north (mean = 724 mm) (IRIMO, 2018). This substantial difference in regional variability of precipitation is most noticeable between the country's north, northwest, west, and central regions (from < 400 to > 2000 mm). In Iran's mountainous regions, altitude has a significant effect on the rainfall, and runoff hydrographs show quite distinct spatial patterns (IEM, 2018).

2.2. TUW model

The TUW model was developed by Viglione and Parajka (2019) as a semi-distributed version of the HBV model (Bergström, 1976). It consists of three routines: snow, soil moisture, and flow response and routing with 15 parameters (Ceola *et al.*, 2015, Parajka *et al.*, 2007). The model treats the elevation zones as discrete entities that contribute to the total output flow in their own right. Daily precipitation, air temperature, and potential evapotranspiration are used as inputs (Fig. 1). Finally, based on the sub-catchment areas, the different outputs from the elevation zones are averaged (Neri et al., 2020). Parajka et al. (2007) and Ceola et al. (2015), respectively, provide more details on the model structure and use in R.

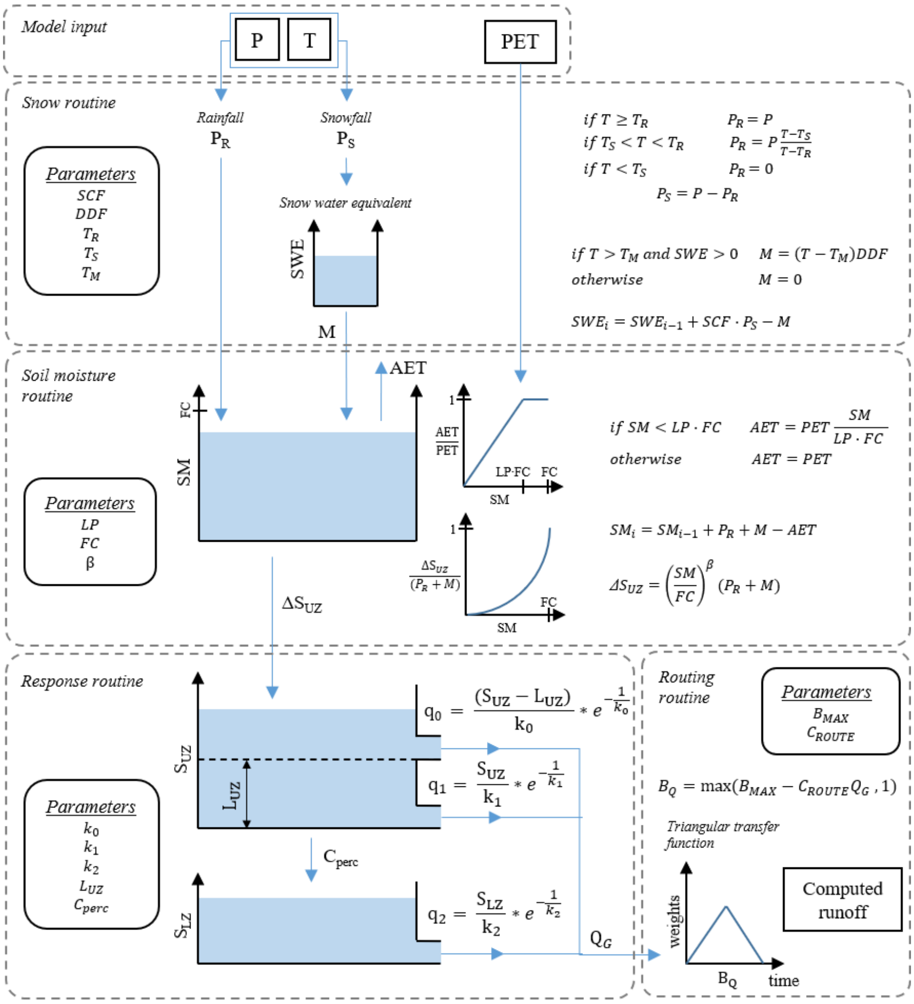

Fig. 1. The structure of TUW model (from Neri et al., 2020).

2.3. Forcing data

Data from the Iran Meteorological Organization (IRIMO) (IRIMO, 2018) and the Iran Energy Ministry (IEM) IEM, 2018) is used to create a daily precipitation time series for all catchments. the IDW and Elevation (IDEW) technique was used to estimate rainfall fields from point measurements measured at gauge locations in this dataset. The IDEW is an interpolation approach that also allows for the possibility of defining elevation weighting along with the distance weighting, making it more suitable for mountainous regions of Iran where topographic influences on precipitation are important. More details are presented in Jahanshahi et al. (2021) and Masih et al. (2011) and Masih et al. (2010)

A regression-based approach and elevation as an explanatory variable was used to construct daily temperature time series from IRIMO/IEM data. The Hargreaves method is used to calculate the reference evapotranspiration based on maximum, minimum, and average temperatures (Hargreaves et al., 1985).

Missing values in the data sets were estimated using the regression method, which relies on values from nearby gauges. To calculate the missed records, the temperature data from nearby gauges were correlated well enough to be used ($R^2 > 0.89$). In the case of precipitation data, the correlation is $R^2 > 0.85$. For all 576 catchments, 6.3% and 9.4% of the temperature and precipitation data, respectively, required to be filled.

As depicted in Figure 2, the climatic variables of the 576 study catchments were assessed between 1993 and 2012. Precipitation, temperature, PET, and snow cover percentage in the study catchments are all subject to some variations between the calibration and validation periods (the validation period is slightly drier than the calibration period). The annual mean of four climate variables in the research catchments, analyzed using Hubert's segmentation approach, shows no trend or change point from

1993 to 2012. Thus, the four variables have inter-annual fluctuation. Annual mean values of climate variables for four climate regions are shown in Table 1 for the years 1993-2012.

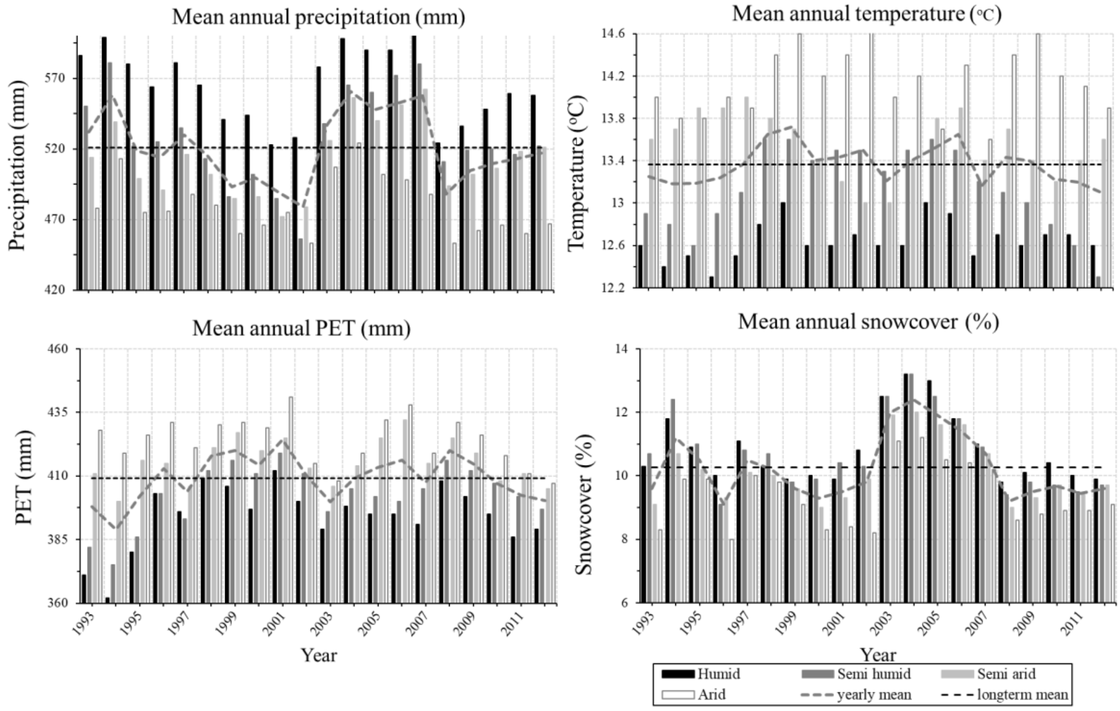

Fig. 2. Climate variability in the study catchments over the period 1993 to 2012.

Table 1

Mean annual values of catchment attributes for four climate regions over two calibration and validation periods.

Region	Calibration and validation periods	Precipitation (mm)	Temperature (ºC)	PET (mm)	Snow cover (%)
Humid	1993-2008	568	12.6	394	10.8
	2008-2012	550	12.6	393	10.1
Semi-humid	1993-2008	527	13.1	402	10.7
	2008-2012	519	13.6	404	9.7
Semi-arid	1993-2008	512	13.5	415	10.1
	2008-2012	511	13.4	410	9.5
Arid	1993-2008	479	14.1	424	9.3
	2008-2012	463	14.2	415	9
Iran (all regions)	1993-2008	523	13.4	409	10.4
	2008-2012	511	13.2	405	9.6

2.4. Catchment dataset

A set of 576 unregulated catchments with a total area of about 407,000 km^2 are selected for this study. These are the same catchments that used in our previous study (Jahanshahi et al., 2021). The catchment area ranging from 64.7 km^2 to 8,432 km^2 and a median of 496 km^2. There is a decline in the catchment area as one moves from humid to arid environments, with the lowest and highest median values occurring in the humid and arid regions, respectively (Table 2). The distribution of meteorological and hydrometric measuring gauges (number and distribution per unit area) worsens from wet to dry.

Land use digital maps (MODIS Land Cover Product), aquifers map (Iran Energy Ministry map), global soil map (based on the FAO map), and the major geological formations (1:250000 map of USGS) are used. These digitized maps are merged with catchment boundaries to determine soil type, land-use type, aquifer area, and geological unit.

The study catchments have a continuous daily streamflow time series from water year (WY) 1992 to 2012 (i.e., September 22, 1992, to September 21, 2012). All discharge data for all catchments are carefully screened and outliers are removed. Table 1 summarizes the median values of CDs for all 576 study catchments. There are correlations between CDs for 576 study catchments, but most of these correlations are not strong. A heatmap of the spearman rank correlations among CDs is presented in Fig. 3.

We split the timeline from WY 1993 to 2012 into the following two calibration and validation periods: WY 1993-2008 is calibration period and WY 2008-2012 is validation period. WY 1992-1993 is used for model warm-up. Figure 4 depicts the locations and classifications of 576 study catchments into four climate regions. For each climate region, the median CD values are shown in Table 2.

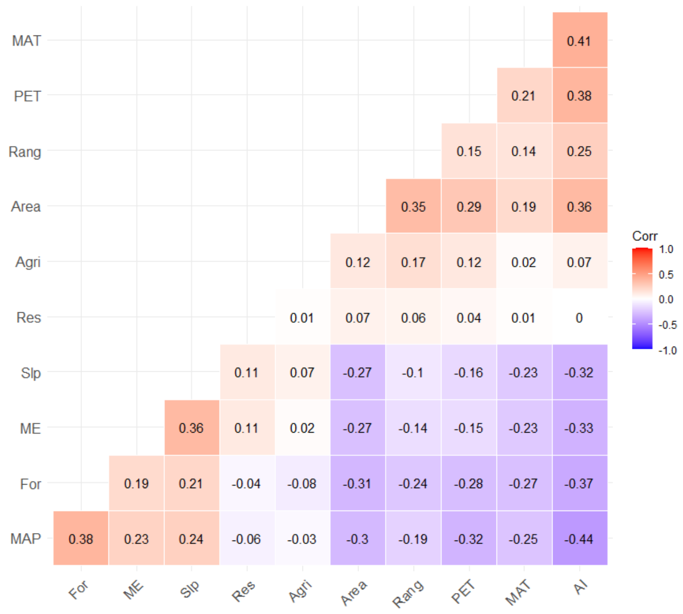

Fig. 3. Heatmap of the spearman rank correlations among catchment descriptors (CDs). AI: aridity index, MAT: mean annual temperature, PET: potential evapotranspiration, Rang: rangeland area, Area: catchment area, Agri: Agriculture area, Res: residential area, Slp: slope, ME: mean elevation, For: forest area, and MAP: mean annual precipitation.

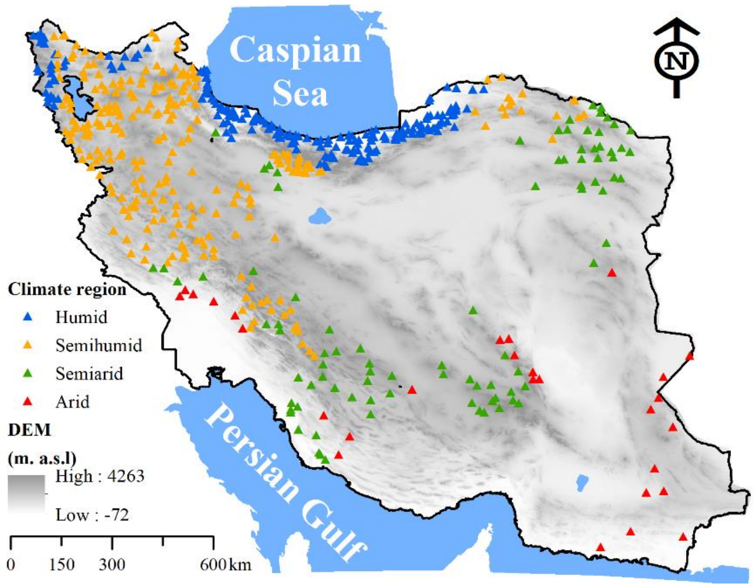

Fig. 4. The location of catchments outlet.

Table 2

The median value of each climate region's catchment descriptors (n = 576).

Catchment descriptor	Humid	Semi-humid	Semi-arid	Arid
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 (^{o}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

3. Methodology

3.1. Strategies

To begin, the best donor catchment, our baseline scenario, is selected based on the highest NSE value after each catchment is considered in turn as an ungauged/target catchment (out of 576 catchments), while the other catchments are considered a donor (576-1 = 575 catchments) (best-donor case or behavioral similarity), and then the effectiveness of two following strategies termed as apparent similarity in prediction of ungauged catchments is evaluated by comparing them to the best-donor case. This is the hypothesis that is used in the study to show the differences between parameter transfer methods to clarify the degree of hydrological similarity. The following are two strategies: Strategy 1; using physical similarity (PS) method (Tables 3 and 4; scenarios 1 through 20) to select physically similar catchments in terms of 11 selected CDs (see Table 3). Potential donor catchments are ranked by PS for each ungauged catchment based on the following: (a) each individual CD rankings and (b) the sum of the 11 CDs rankings. Strategy 2; using the nearest neighbor (NN) method to prioritize donor catchments for model parameter transferring to target catchments (Table 5, experiments 1 through 4). The geographical distance between gauged and ungauged catchments is used in this strategy.

Table 3

Median spatial (km) and Euclidean distance (-) between gauged and ungauged catchment pairs for spatial proximity and physical similarity methods (n = 576).

Regionalization approach	Scenario No.	Median spatial and Euclidean distance between gauged and ungauged catchment pairs
SP (i.e., nearest neighbor method)	-	71.2
PS (all eleven CDs)	S1	689.5
PS (aridity index)	S2	422.4
PS (mean annual precipitation)	S3	441.1
PS (mean annual temperature)	S4	460.9
PS (PET)	S5	485.4
PS (area)	S6	509.7
PS (mean elevation)	S7	588.9
PS (mean slope)	S8	617.1
PS (all three topographic and physiographic CDs)	S9	575.7
PS (rangeland)	S10	847
PS (agriculture)	S11	880.3
PS (forest)	S12	1,077.9
PS (residential)	S13	1,232.8
PS (all four land use CDs)	S14	1,007.5

Note: SP is spatial proximity, PS is physical similarity, and CDs is catchments descriptors.

Table 4

Median Euclidean distance (-) between gauged and ungauged catchment pairs for physical similarity method, stratified by spatial scales (local and regional).

Region	Scenario No.	Catchment descriptor to calculate the SI	Runoff ratio	Median Euclidean distance between gauged and ungauged catchment pairs
Iran/local (n = 576)	S15	Aridity Index (-) and mean elevation (m)	0.11-0.78	481.3
	S16	Aridity Index (-) and area (m)		493.2
Humid (n = 199)	S17	Aridity Index (-), area (km^2) and mean elevation (m)	0.17-0.78	479.5
Semi-humid (n = 256)	S18	Aridity Index (-), area (km^2), mean elevation (m) and mean slope (%)		509.12
Semi-arid (n = 93)	S19	Aridity Index (-), area (km^2) and agriculture (km^2)	0.11-0.61	572.6
Arid (n = 28)	S20	Aridity Index (-), area (km^2), agriculture (km^2) and rangeland (km^2)		583.4

Table 5

All examined experiments for nearest neighbor transfer method (n = 576).

Row	No. of experiment	Geographical distance from donor catchment	No. of donor catchments
1	E1	0-20	34 (6%)
2	E2	0-40	118 (20%)
3	E3	0-60	241 (42%)
4	E4	0-90	362 (63%)
5	E5	0-10	14 (2%)
6	E6	10-20	20 (3%)
7	E7	20-30	39 (7%)
8	E8	30-40	45 (8%)
9	E9	0-68	289 (50%)

3.2. Evaluate the consistency of measured hydrological and physical similarities

We examine Strategy 1 in this section. We assess a methodology to evaluate the consistency of the measured hydrological and physical similarities for regionalization, based on the assumptions described by Oudin et al. (2010). Here, we first establish metric that measure physical and hydrological similarities.

3.2.1. Measuring the hydrological similarity

To define the HS, we compared the outlines of two types of models: (1) a model with ~~regionally or~~ locally calibrated parameters (gauged catchment) and (2) a model with parameters estimated by calibration on another (donor) to target (ungauged) catchment in terms of (i) catchment descriptors (PS) and (ii) the highest NSE values in parameter transfer, regardless CDs (best donor case (BD)). Parameter transferability can be used to define HS (Oudin et al., 2010). Transferring whole parameter sets are preferred to the individual parameters in order to eliminate interactions between parameters (McIntyre et al., 2005). The ability of parameter sets to simulate streamflow can be taken into account to select hydrologically similar catchments. To this end, if the model's efficiency on ungauged catchment obtained by the model using the parameters estimated by calibration on gauged catchment is greater than 0.7 of model's efficiency estimated in the BD case on ungauged catchment, the gauged catchment is considered as hydrologically similar to ungauged catchment (this is our study assumption) (section 4.3). The value of 0.7 is appointed based on the difference between calibration results and PS achieved after the initial implementation of the parameter transfers. Thus, our assumption is that this distinction was examined in two ways: (1) if the difference between calibration and PS is less than 30% and (2) if the difference in performance between PS and BD is less than 30%, the performance of hydrologically (behavioral) similar catchments is considered "good" (Oudin et al., 2010 adopted the value of 10% in their study).

3.2.2. Measuring the physical similarity

We employ the methodology by Kay et al. (2007) to measure the PS between catchments, in which Euclidean distance is used to define the similarity in CDs space, and CD values are normalized by their standard deviation over the entire catchment set. Here, we considered the following weighting of CDs:

$$dist_{a,b} = \sqrt{\sum_{j=1,J} w_j \left(\frac{X_{a,j} - X_{b,j}}{\sigma_{X,j}}\right)^2} \tag{1}$$

where j indicates one of a total of J CDs, $X_{a,j}$ is the value of that CD at the ath catchment, $\sigma_{X,j}$ is the standard deviation of the CD across the entire catchment set, and w_j is the weight attributed to the jth CD. To begin, distances involving only one CD were tested (i.e., by setting the weights of the other descriptors to zero). Following that, weights were considered to be identical for all CDs, as is common in PUB studies (e.g., Oudin et al., 2008; Parajka et al., 2005). Last, the weights were optimized to maximize the overlap between physically similar catchments and hydrologically similar catchments (Oudin et al., 2010).

For the optimization, more than 3000 weight combinations were tested, reflecting every possible combination, with weights ranging from zero to unity with an increment of 0.1 and the sum of the weights being equal to unity. We assessed the relevance of eleven physical similarity measures based on equation (1) to evaluate the consistency of physical and hydrological similarity. For all CDs, all weights are set to zero except one; that is, only one CD is used to assess the physical similarity between the target catchment and the other catchments. This set of combinations tests the relevance of each CD in determining HS. Then, we employed multiple CDs simultaneously in equation 1 to find the set of CDs that produced the highest degree of HS. By repeating the process, we eventually identified the scenarios (Table 4).

Three sets of catchments are determined here as follows:

(1) a group of catchments that are the most physically similar to each catchment considered in turn (consider physical cousin by Oudin et al., 2010). This group is determined using distance calculation (equation 1). The number of catchments in this group is n = 478.

(2) those catchments that produce the highest NSE values in BD are likely to be hydrologically similar catchments (hydrological cousin). The number of catchments in this group is n = 437. The overlap between these two sets indicates the ability of the CDs to identify the catchments that are hydrologically similar (our assumption).

(3) a group of catchments that are the most spatially similar to each catchment considered in turn (spatial cousin). The number of catchments in this group is n = 494.

It's worth mentioning that we incorporate the NN method into the Oudin et al. (2010) methodology. To do so, we employed both PS and NN to calculate the HS, allowing us to compare the two methods more closely. The PS considers individual CDs, the NN considers different GD between donor and target catchments, and the optimum option for each is then compared to the calibration and BD cases (compare three approaches to achieve the best solution for model parameter transfer). A detailed stepwise implementation of all processes and parameter transfers strategies are presented in Fig. 5.

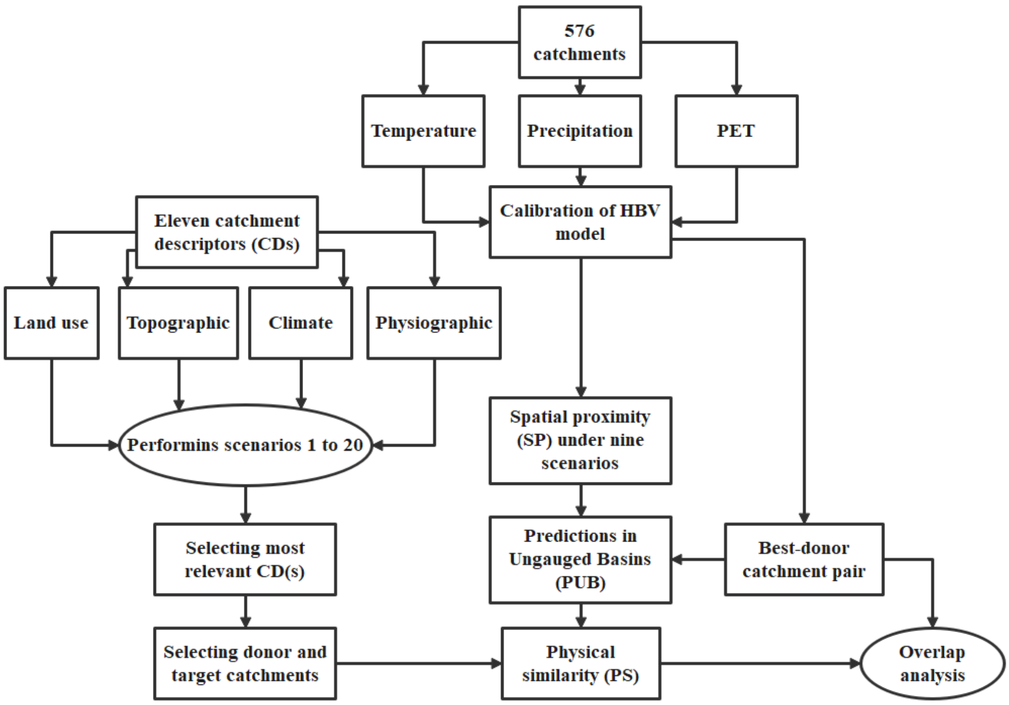

Fig. 5. Flowchart representing all processes performed in this study.

4. Results

4.1. Model performance for calibration period (at site)

We used from calibrated parameters estimated in our previous study (Jahanshahi et al., 2021). The median NSE values of calibration period (1993-2008) for all 576 gauged were greater than 0.5. The results indicate that the model performs better in wetter catchments than it does in drier ones. These findings generally corroborated those of Oudin et al. (2008) in France and Parajka et al. (2005) in Austria.

In Fig. 6, the 576 catchments' model performance is shown as a regional distribution. According to the model's performance in the northwestern, northern and inner western regions of the country, they are far superior than the other regions. Rainfall patterns in interior, western, and southeast catchments vary in amplitude, making streamflow modeling more complex. When it comes to model performance, highland catchments typically outperform lowland ones.

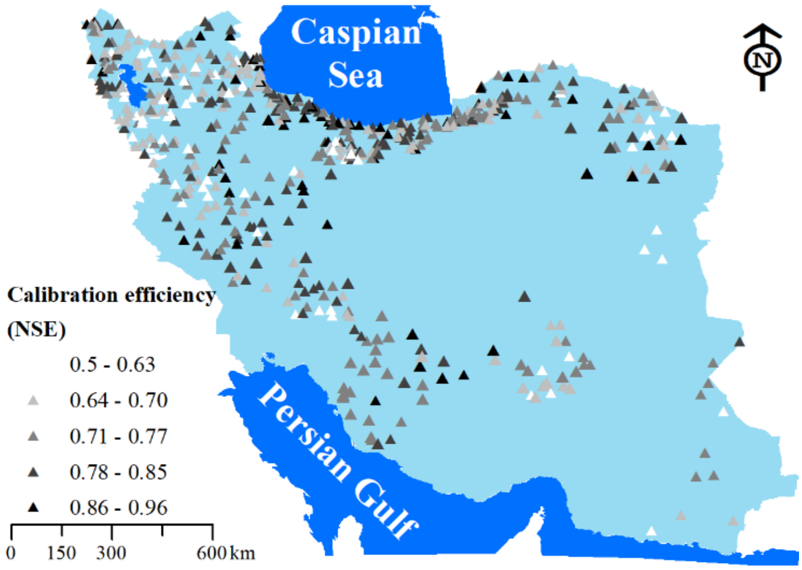

Fig. 6. Calibration performance of the model over the 576 study catchments.

4.2. Relating similarity index with transfer performance throughout Iran (local) and climate regions (regional)

We investigate the effect of similarity in catchment descriptors and geographical distance on successful parameter transfer (hydrological similarity). Transferring the best parameter sets between 576 catchments results in 993,600 (575 for three approaches) cases. About 88.7% of the 993,600 cases, produced an NSE greater than zero. The physical similarity and nearest neighbor methods are compared to identify the main controls on parameter transfer at two different scales: climate regions (regional) and the entire Iran (local). To select the most appropriate approach and quantify its performances, the calibration and best-donor cases are used as the base-case and second-base, respectively. Here, the PS method employed eleven CDs to define catchment similarity. When it came to hydrological indices computed in PUB studies, the most variability was explained by these CDs (e.g., Arsenault and Brissette, 2014; He et al., 2011; Petheram et al., 2009; SKM, 2009; X. Yang et al., 2020; Yang et al., 2018). The median distance between gauged and ungauged catchment pairs is shown in Table 3.

To test the relevance of each CD in determining hydrologic similarity, the relevance of eleven PS measures is assessed based on Euclidean distance equation (Section 3.2.2). In ten of them, all weights are set to zero except one, implying that only one CD is used to assess the PS between target/ungauged catchment and the other catchments. This set of combination tests examines the utility of each CD in determining hydrological similarity. One such combination is a PS measure that assigns equal weights for all CDs. Therefore, potential donor catchments are ranked by PS for each target catchment based on: (1) one CD, and (2) a combination of eleven CDs (Table 3). When only one CD is used in the similarity equation, four appear to be more relevant than the others: the aridity index (AI), mean elevation (ME), catchment area (AR), agriculture cover (AGR), and rangeland cover (Rang). Then comes the PET. Forest cover and residential cover are the two CDs that have the least relevant on the selection of hydrologically similar catchments. The low performance of similarity in these two CDs, could possibly be explained by the more variable forest/residential characteristics identified in study catchments; however, this is hypothesis, and more research is needed to discover why forest and residential are ineffective as useful characteristics.

The median distance between donor and target catchment pairs in terms of both CDs and GD groups is shown in Table 4. SI values for sum of the five-best CDs are also shown in Table 4 at two different scales: of local and regional. Sixteen scenarios are examined to ascertain the dominant controls on successful parameter transfer at two scales (see Tables 3 and 4). Figure 7 illustrates the box-plot comparison of the ten-best PS scenarios, as well as the performance of the NN and BD cases in calibration and validation modes. As seen in this Fig, the best performance for PS is achieved with S15 (median NSE = 0.56) (in which AI and ME are employed to select most similar donor target catchment pairs), followed by S2 (decline of 17.4% (compared to calibration)), S6 (decline of 19%), S7 (decline of 22.9%), S3 (decline of 24.5%), S5 (decline of 28.5%), S4 (decline of 30.3%), S9 (decline of 34.3%), S11 (decline of 39.7%), and S20 (decline of 43.8%). In general, we observed that scenarios based on climatic descriptors (aridity index) performed better in the PS than scenarios based on topographic (ME), physiographic (catchment area), and land use (catchment percent agriculture) descriptors. ME is chosen as the most prominent topographic descriptor on parameter transfer, which is also considered the second dominant control. Hence, generally, our findings indicated that when we identified the similarity in climate and ME (particularly at elevations), the results resulted in a more successful parameter transfer than with other CDs at the regional scale. Among four land use descriptors examined in PS (Table 3), the use of catchment percent agricultural and rangeland in define similarity, resulted in improved parameter transfer performance.

The performance ranking of CDs in validation mode is similar to that in calibration mode, although the difference in performance is significant (scenarios 15 and 20 are selected as the best and worst, respectively). This finding reveals that moving between periods (calibration to validation) considerably impairs the performance of both transfer methods. This conclusion is consistent with PUB studies (e.g., Jahanshahi et al., 2022, 2021; Oudin et al., 2008; Patil and Stieglitz, 2015; Petheram et al., 2009; Zhang and Chiew, 2009).

As expected, the best strategy is BD (median NSE = 0.65). Its difference from S15 (the best scenario in PS) is 6.62% for calibration and 2.53% for validation. The NN (median NSE = 0.5) is chosen as the third-worst strategy. This results demonstrated that the GD has a more influence on parameter transfer than temporal gap. This effect was significantly enhanced in validation mode, significantly impairing the NN performance. As a result, the NN performance is reduced by 32.12% when compared to BD performance (reduces 23.46% for calibration mode). This overall finding is consistence with Patil and Stieglitz (2015) at 294 catchments across US. The main reason for the relatively poor performance of the NN is that the median distance between catchment centroids is relatively large (greater than 71 km), which means that selecting donor-target catchment pairs on the basis of SP method will provide no better results than using the PS and BD methods. The climatic differences between the calibration and validation periods could also contribute to performance decline from calibration to validation. It is a little wetter during the calibration (1993-2008) period than during the validation (2008-2012) period (see Table 1).

The box-plots of controls on parameter transfer for wet (humid and semi-humid) and dry (arid and semi-arid) catchments are shown in Fig. 8. For wet catchments in calibration mode, S16 (where aridity index and catchment area are used to select most similar donor target catchment pairs) produced the best results, followed by S7 (decline of 17.9%) (compared to calibration)), S2 (decline of 19.6%), S6 (decline of 22.6%), S3 (decline of 26.7%), S9 (decline of 30.8%), S20 (decline of 32.4%), S5 (decline of 36.4%), S4 (decline of 36.6%), S11 (decline of 42.8%). The CDs performance ranking is kept in validation mode just like it is in calibration mode. The best and worst results for arid and semi-arid regions are obtained using S19 and S20, respectively.

The spatial distribution of catchments where five-best PS scenarios at both regional and local scales are shown in Fig. 9 (left). We found that similarity in CDs from three categories - climate (AI), topographic (Me and slope), and physiographic (area) are required for successful parameter transfers in humid and semi-humid regions (455 parameter transfers) (the very northwestern and northern humid catchments as well as northern coastline at southern fringe of Caspian Sea). Similarity in agriculture (land use) is added to those for two aforementioned climate regions for successful parameter transfer in arid and semi-arid regions (121 parameter transfers) (the northeastern and middle mountains as well as middle and southeastern plains). As a result, for arid region, the only land use descriptor used to judge the effectiveness of a parameter transfer for 1.04% of catchments (n = 6) is similarity in catchment percent agriculture (S19). This means that land use descriptors do not have as much of an impact on parameter transfer as climatic, topographic, and physiographic descriptors, which have the most potential for determining parameter transfer similarities. The results of Singh et al. (2014) in 83 catchments across the US show a high potential for successful parameter transfer based on similarities in climate and topographical characteristics. They found that similarity in percentage of agriculture has the most potential for parameter transfer for humid and semi-humid catchments of the plains. Our overall findings do not support this conclusion. Similarity in elevation was found to be the most important control on parameter transfer for humid and semi-humid plateaus in their study, which is consistent with our high-altitude catchments in humid and semi-humid regions. PET and ME are the most relevant controls on a successful parameter transfer in semi-arid mountains and plateaus across the US, whereas, area, aridity index, and ME have the largest impact in our study. Beyond this, other CDs do not provide any more useful information. In semi-arid region (S20), similarity in percentage of rangeland was only able to successfully transfer parameters for one catchment. We also considered the role of the runoff ratio in determining whether or not a parameters transfer was successful (not in similarity for PS). The runoff ratio values were found for humid and semi-humid regions, where parameter transfer is most facilitated by similarity in climatic and topographic characteristics. In contrast, lower runoff ratio values for arid and semi-arid regions (drier catchments) demonstrates how all four CD types come together to determine the successful of parameters transfer (Table 4, S19 and S20). Figure 9 (right) illustrates the performance of BD in catchments classified as: (1) 0.4-0.6, (2) 0.6-0.8 and (3) $\geq$ 0.8. By comparing the catchment distribution in the right and left panels of Fig. 9, we found that the strong performance of S15, S2 and S7 is consistent with the performance of the BD case, particularly in humid and semi-humid regions. This finding revealed that similarity in aridity index and mean elevation plays a significant role in determining hydrological similarity for these wet catchments. The S6 and S19 performs relatively well in terms of compliance with two moderate categories of BD (0.4-0.6 and 0.6-0.8) when compared to other scenarios for drier catchments of semi-arid and arid regions. This finding confirms that the catchment area and agriculture cover are both high relevant CDs in the selection of hydrologically similar catchments among these two groups' catchments.

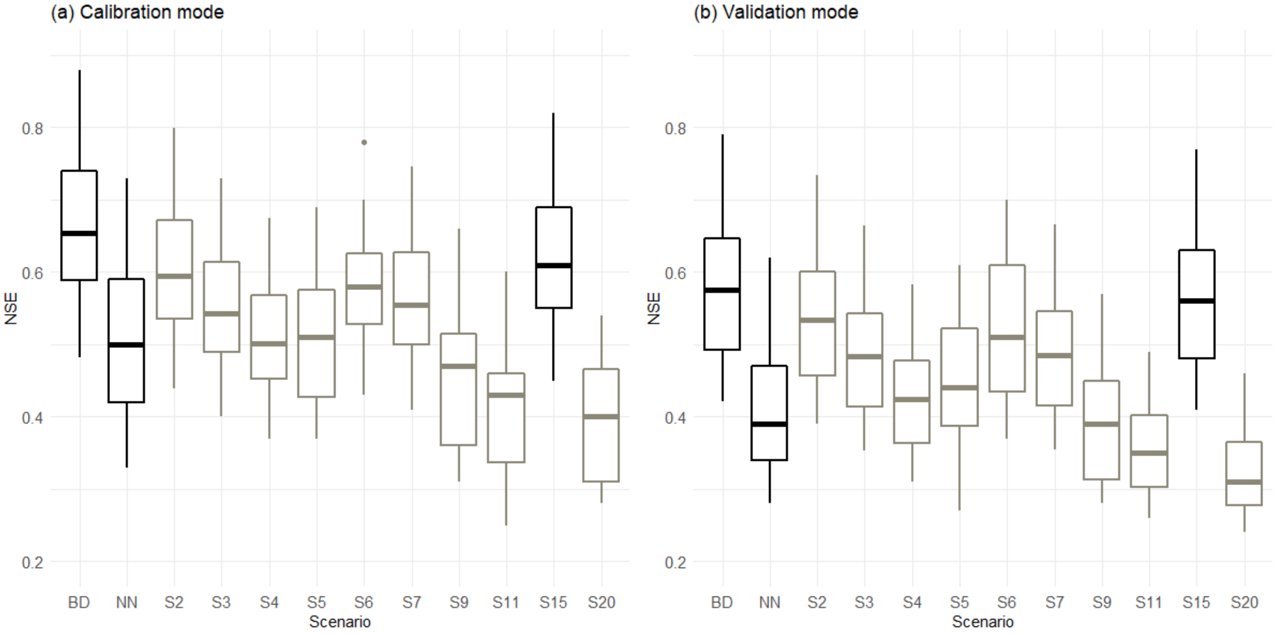

Fig. 7. A box-plot comparison of nearest neighbor (NN), best-donor (BD), and ten-best parameters transfer scenarios in physical similarity for (a) calibration and (b) validation modes. Tables 3 and 4 contain descriptions of all scenarios.

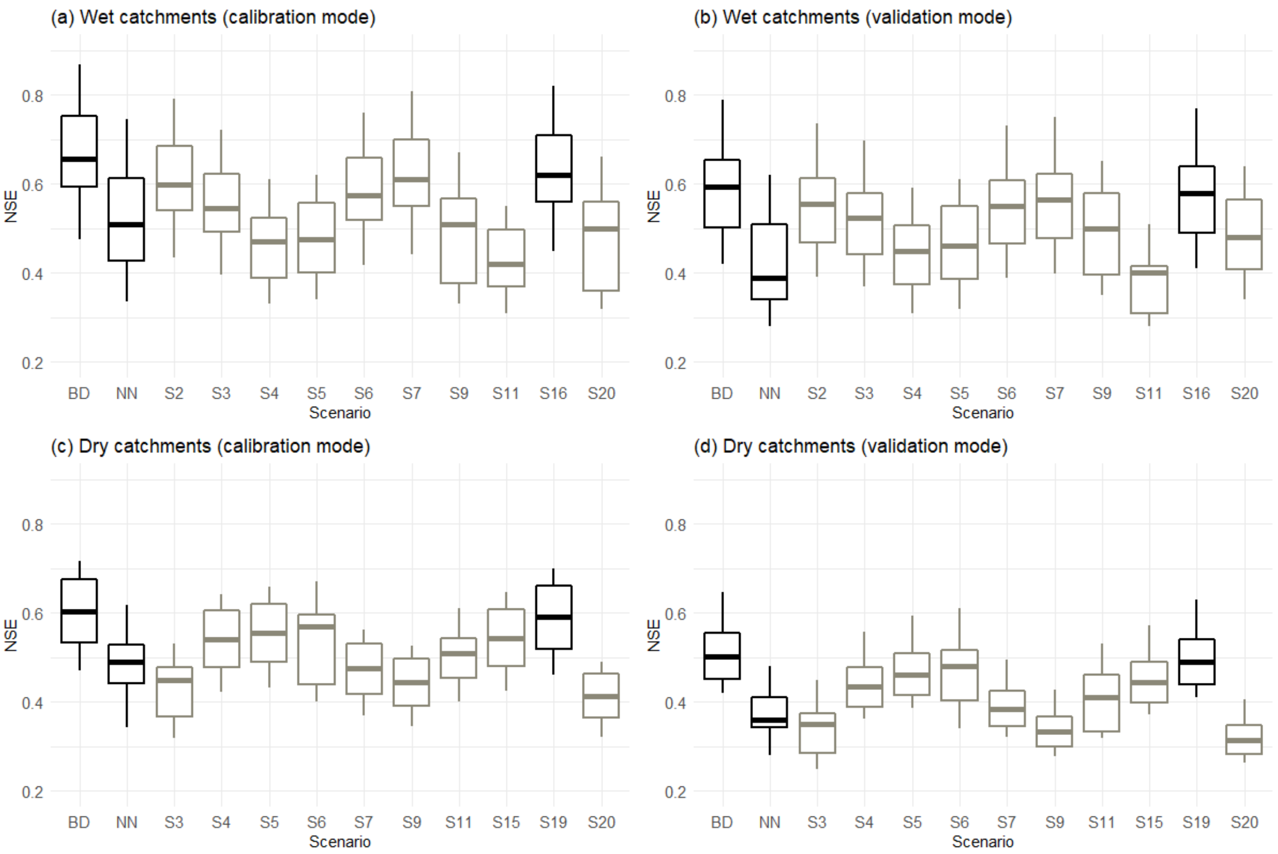

Fig. 8. A box-plot comparison of the nearest neighbor (NN), best-donor (BD), and ten-best parameters transfer scenarios in physical similarity for wet and dry catchments; (a) and (c) calibration, (b) and (d) validation modes. Tables 3 and 4 contain descriptions of all scenarios.

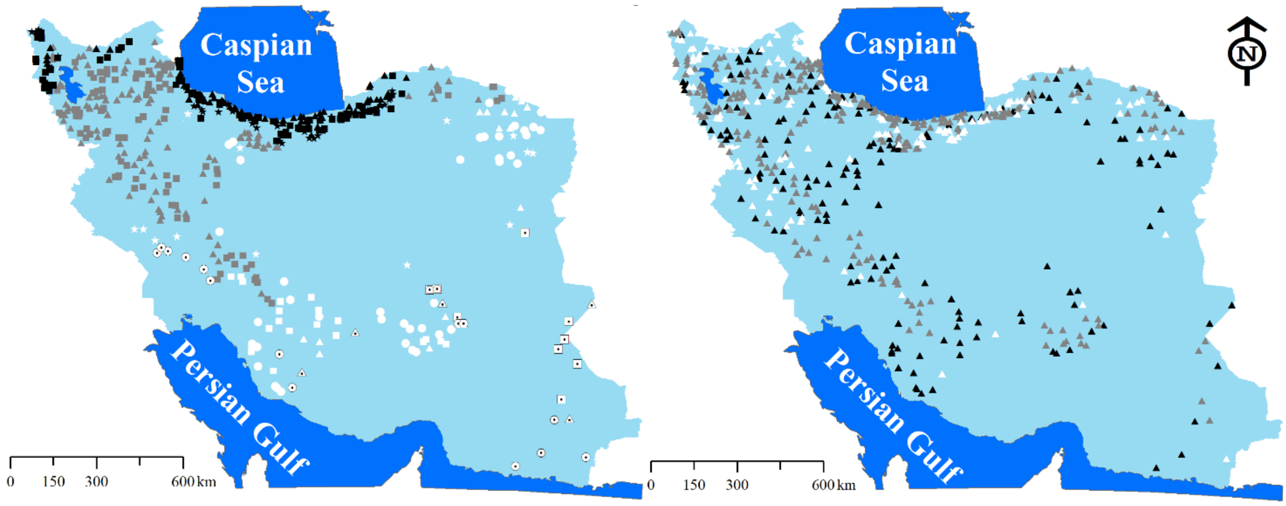

Fig. 9. Location of 576 catchments where; (left) S15 (n = 203; 35.2%) (square), S2 (n = 201; 34.5%) (triangle), S6 (n = 70; 12.5%) (circle), S7 (n = 96; 16.6%) (star), and S19 (n = 6; 1.04%) (hexagon) perform best in PS for humid (black), semi-humid (grey), semi-arid (white), and arid (white with a black dot); (right) median NSE value in BD case is: 0.4-0.6 (n = 176; 3%) (black triangles), 0.6-0.8 (n = 268; 46%) (grey triangles) and ≥ 0.8 (n = 132; 23%) (white triangles) in the calibration mode at regional scale.

4.3. Evaluating the overlap in physical similarity by blending the catchment descriptors

The selection of hydrologically similar catchments improves remarkably when the CDs are blended by using their combinations (Tables 3 and 4). As a result, the S15 combination had the best overall performance throughout Iran. To better understand the role of PS in HS performance, we investigate (i) the difference between calibration (best-case) and PS (S15) (Fig. 10; left) and (ii) the overlap between PS and BD (second-best case) (Fig. 10; right). We organized the catchments into four groups based on this overlap: (1) ≥ 90, (2) 70-90, (3) 50-70, and ≤ 50. The results showed that, when these categories are considered, the overlap is greater than 70% for more than 75% (n = 431) of the catchments, implying that (1) physically (apparently) similar catchments have a high similar hydrological behavior and (2) the CDs in S15 (aridity index and ME) are able to define hydrological (behavioral) similarity well on the basis of transferability of model parameters (Fig. 10; right). Out of these 431 catchments, 307 (71%) were wet, and 124 (29%) were dry. We also found that the smaller gap between calibration and PS is associated with wetter (high-altitude) catchments in humid and semi-humid regions, where the aridity index and ME are most relevant controls in selecting hydrologically similar catchments (Fig. 10; left). Two possible reasons for reduction the overlap (≤ 50%) for 9% (n = 53) of catchments are: (1) the lack of CDs that would potentially increase overlap between two hydrological and similar cousins (e.g. groundwater and geologic) and (2) the lack of hydrological cousins for some ungauged catchments. Most of these catchments are located in semi-arid and arid (very southwestern, middle, and southeastern plains) as well as semi-humid (middle western and eastern catchments) regions which have drier climatic conditions.

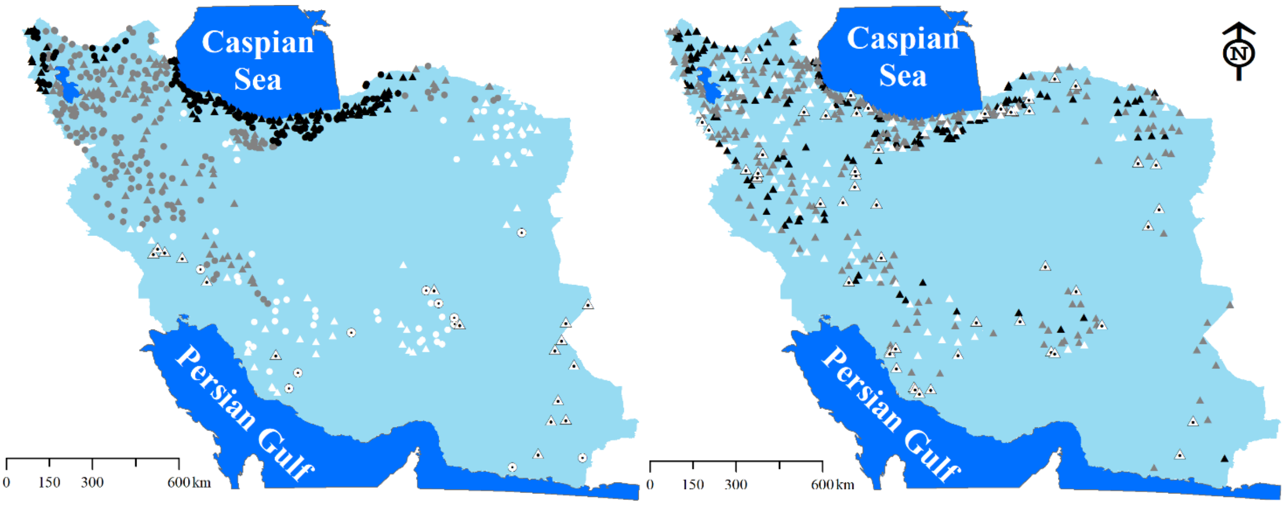

Fig. 10. Location of 576 catchments where; (left) the difference between median NSE value in calibration and PS (S15) is; ≤ 30% (n = 327; 56.7%) (circle), ≥ 30% (n = 249; 43.2%) (triangle) for humid (black), semi-humid (grey), semi-arid (white), and arid (white with a black dot) at the local scale; (right) the overlap percentage between PS and BD is; ≥ 90 (26%; n = 149) (black triangles), 70-90 (49%; n = 282) (grey triangles), 50-70 (16%; n = 92) (white triangles), and ≤ 50 (9%; n = 53) (white triangles with dot) at the local scale (calibration mode).

4.4. Evaluation of spatial proximity

To calculate catchment similarity, the GD between the donor and target catchment is employed as an attribute in the NN approach. Because of the wide range of climate heterogeneity in Iran, it is necessary to determine the optimal distance between the centroids of donor-target catchment pairs in order to maximize the spatial transfer of parameter sets in order to acquire the most accurate results. The NN is used in four experiments to achieve this goal. (a) distance (d) less than 20 km, (b) $d \leq 40$ km, (c) $d \leq 60$ km, and (d) $d \leq 90$ km are defined as donor-target catchment pairs in these experiments. Results of these experiments are summarized in Table 5 (rows 1 to 4). Figure 11 shows the NSEs of these experiments in box plots.

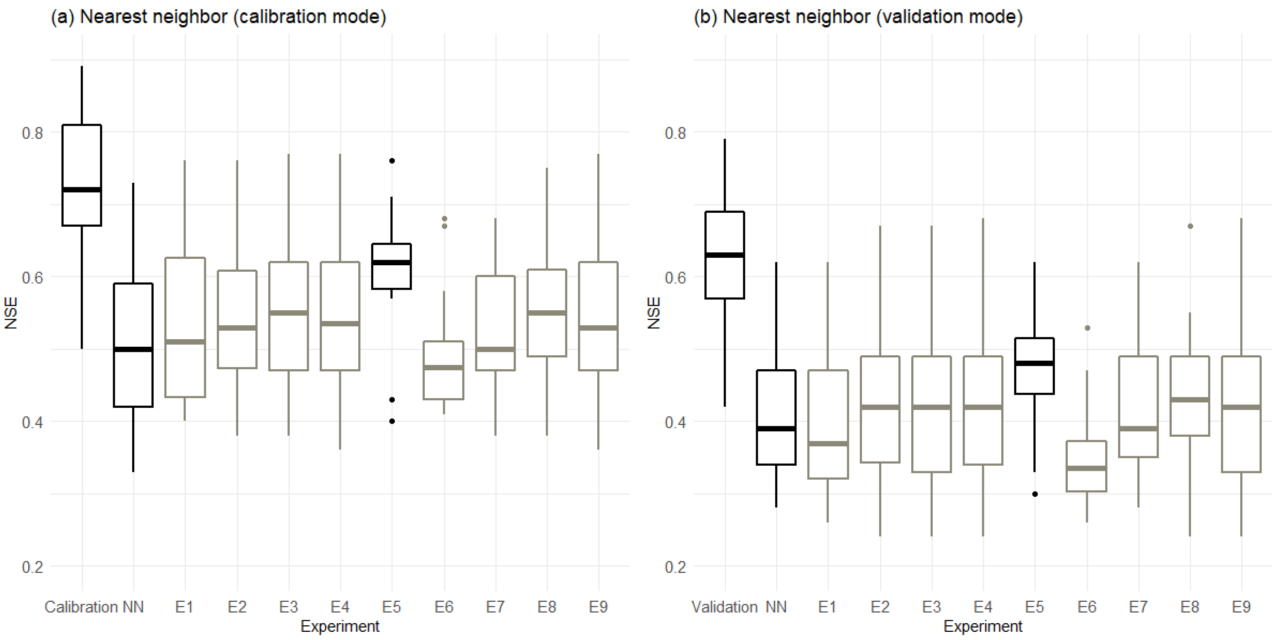

Fig. 11. A box-plot comparison of the nearest neighbor method's tested experiments for (a) calibration and (b) validation modes. NN is nearest neighbor. Table 5 contains descriptions of all experiments.

In both calibration and validation modes, as shown in Fig. 11, significant differences occur between the results of experiments and the NN method. When compared to NN, the performance of E1 (0 to 20 km) increased by 25.37% and 32.75% for calibration and validation modes, respectively. Table 5 shows that more than half of the study catchments (about 63%; n = 362) are within a distance of 0 to 90 km. The GD was subsequently broken into five smaller spatial distances, and the NN was re-implement under them. These new spatial distances are as follows: (1) 0-10 km, (2) 10-20 km, (3) 20-30 km, (4) 30-40 km and (5) 0-68 km (Table 5 rows 5 to 9). Our goal is to conduct a more precise investigation of the optimal GD in the NN method, then compare the optimal results to the performance-related modifications of blending the CDs in defining similarity for PS. Comparing PS and NN reveals that, while the NN performs worse than the PS, changing the GD between the donor-target catchment pairs has a more complex effect on performance than CD combinations (no further details are provided), and the optimal GD for model parameter transfer lies between 0 and 10 km (n = 14).

5. Discussion

We find interesting differences in physical (physical similarity method) and spatial (nearest neighbor method) controls on model parameter transfer when investigating our dataset spanning Iran (local) and when dividing it into climate regions (regional). To investigate the CDs-geographical distance-performance link to transfer parameter sets between the 576 unregulated catchments, we apply three strategies: PS, NN, and best-donor (best donor with highest NSE value). Previous PUB studies in Iran have not directly made this comparison. Parameters transfer was successful when employing NSE, as it was in PUB studies (Chouaib et al., 2018; Merz and Blöschl, 2004; Oudin et al., 2010, 2008; Parajka et al., 2005; Zhang and Chiew, 2009; Jahanshahi et al., 2021). Although the main aim of PUB studies was to simulate in ungauged catchment, the success of these simulations was not attributed to the catchment descriptors input combinations.

For PS, the main controls for parameter transfer throughout Iran (the local scale) are climate and topographic, followed by physiographic. Other PUB studies have found that the importance of (i) climatic gradients (Chouaib et al., 2018; Sawicz et al., 2014), (ii) topographic (Price, 2011), and (iii) land use form, climate descriptors, and geologic structure (Winter, 2001; Wolock et al., 2004) across the USA. Our results agree with their finding, though we also found that climate, physiographic, and land use are major controls on parameter transfer at regional (semi-arid and arid regions, where that catchments have similar energy conditions and are water-limited) as well as local scales. According to Singh et al. (2014), climate is not the most relevant control on parameter transfer in humid regions across the US, but physical attributes and land used patterns are most dominant. Their conclusion contradicts our findings. ME, on the other hand, was chosen as an important secondary control in mountainous catchments.

In humid and semi-humid regions (where energy-limited conditions are prevalent), the study catchments have high elevation, highest variation on elevation, the highest percentage slope and largest percentage of land covered by forest (see Table 2). We can assume that this highly variable topography is associated with a variable climate gradient. The difference between our study and a previous regional study in 903 catchments in France and UK by Oudin et al. (2010) is that; (1) we did not seek to find the number *n* of hydrologically similar catchments, instead, we compared the most physically similar catchment to the best-donor case to determine the overlap percentage and hydrological similarity accuracy and (2) the study models are different (HBV versus GR4J and TOPMO). Our results demonstrated that more than 75% of our physically similar catchments have a hydrological similarity, with a substantial overlap (more than 70%), whereas only 60% of the hydrologically similar catchments are physically similar.

In general, the following controls appear in sequence when classifying the performance of parameter transfer based on all tested options; (1) spatial distance (for GD $\leq$ 10 km), (2) climate, topographic, physiographic and land use characteristics. Therefore, geographically similar catchments are found to be more hydrologically similar than physically similar catchments only for GD less than 10 km, implying that the PS method provides useful information for spatial distances more than 10 km. In comparison to wetter catchments, our findings imply that physiographic (catchment area) and land use (percentage of agriculture and rangeland area) descriptors are prevalent in drier catchments. The overall performance of model transferability corroborated the general pattern of humid catchments outperforming dry catchments observed in the PUB-reviewed studies of Parajka et al. (2013).

To understand the results more accurately between climate regions, we entered the runoff ratio (streamflow signature) (see Table 4). We concluded that the catchments with higher runoff ratio have better calibration and regionalization performance, as evidenced by 455 wet catchments (79%)

(humid and semi-humid regions) out of 576 study catchments, where the median NSE values for calibration and PS cases range from 0.65 to 0.89 and 0.6 to 0.82, respectively, and the runoff ratio ranges from 0.17 to 0.78 (median = 0.43). The runoff ratio ranges from 0.11 to 0.61 (median = 0.24) for 121 dry catchments (21%) (semi-arid and arid regions) with median NSE values for calibration and PS cases ranging from 0.5 to 0.64 and 0.5 to 0.59, respectively (see Table 4).

Some model regionalization studies have shown that combining streamflow time series from several parameter sets from a multi-donors ensemble generally works better than a single donor framework (Oudin et al., 2008; Viney et al., 2009; X. Yang et al., 2020). However, our goal was to identify only one donor catchment (physically cousin/BD/spatially cousin) most similar to the target catchment. Different results from using multi-donors should be considered in future investigations.

6. Conclusions

In this study in order to explore the links between apparent physical/spatial similarity and hydrological similarity we related (i) physical similarity in several catchment descriptors (CDs) and (ii) geographical distance (GD), to behavioral similarity (the highest-performance in hydrologic model parameter transfer, which is considered the best-donor case (BD)) by analyzing the overlap between them. Our aim is to use a simple strategy to determine the most relevant available CDs or GD to explain the behavioral similarity. On the basis of HBV model parameter transfer to the ungauged catchment, the optimal: (1) GD and (2) CDs were determined in this strategy.

By comparing three pools of hydrologically (hydrological cousins), physically (physical cousins), and spatially (spatial cousins) similar catchments, defined on the basis of available CDs and GD, throughout Iran and subsequently in smaller climate regions, we found that GD is the main control on successful transferability. At the scale of entire Iran (local), by overlap analysis between physical similarity and best-donor case, the geographical distance (less than 10 km) in nearest neighbor, followed by a combination of climate (aridity index) and topographic (mean elevation) descriptors in PS, are selected as dominant controls on parameters transferability to the ungauged catchment. Thus, most relevant CDs guarantee the selection of hydrologically similar catchments for 75% of the ungauged catchments (for which the overlap was more than 70%). By categorizing catchments into climate regions, physiographic (catchment area) and land use (agriculture and rangeland classes) descriptors are added to the most relevant spatial controls. In almost all catchments, all four types of CDs - topographic, climate, physiographic, and land use emerged as important controls on parameter transferability. The runoff ratio (functional characteristic) revealed that higher values are associated with regions where the climate descriptors are main controls. Thus, in general, the shortest GD (less than 10 km) between donor and target catchment emerged as the most prominent control on parameter transferability. This implies that the identifying an appropriate metric of hydrologic similarity depend on the (i) geographical distance and (ii) CD type.

This study had some limitations, which will be fascinating topics for future research, as discussed here. First, due to lack of data, similarity in geology, soil and groundwater characteristics is restricted throughout Iran. Second, the number of catchments available in dry regions (only 28 catchments) were limited. Third, there are some possible sources of uncertainty (although uncertainty analysis was not part of this study's scope) as well as non-stationary climate conditions.

The main advantage of our study is that it is not possible to consider the similarity only in a particular catchment descriptor or geographical distance at Iranian catchments to ensure successful model parameters transfer, but we must consider -a certain combinations- or a wide range of them (both CDs and geographical distance) for different climate regions. Therefore, if we are able to establish that similarity at an optimal geographical distance and with particular characteristics, it would provide an appropriate model transferability. To strengthen the role of physical similarity parameter transferability, further research is needed to identify more relevant lithologic/geologic/groundwater/soil descriptors.

Acknowledgments

The authors appreciate the assistance and data provided by the Iran Meteorological Organization (IRIMO) and the Iran Energy Ministry (IEM). We also thank Lieke A. Melsen for her comments on an early draft of this manuscript.

References

Andréassian, V., Perrin, C., Michel, C., 2004. Impact of imperfect potential evapotranspiration knowledge on the efficiency and parameters of watershed models. J. Hydrol. 286, 19–35. https://doi.org/10.1016/J.JHYDROL.2003.09.030

Andréassian, V., Perrin, C., Michel, C., Usart-Sanchez, I., Lavabre, J., 2001. Impact of imperfect rainfall knowledge on the efficiency and the parameters of watershed models. J. Hydrol. 250, 206–223. https://doi.org/10.1016/S0022-1694(01)00437-1

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

Blöschl, G., 2005. Rainfall-Runoff Modeling of Ungauged Catchments. Encycl. Hydrol. Sci. https://doi.org/10.1002/0470848944.hsa140

Blöschl, G., Sivapalan, M., Wagener, T., Viglione, A., Savenije, H., 2013. Runoff Prediction in Ungauged Basins: Synthesis across Processes, Places and Scales. Runoff Predict. Ungauged Basins Synth. Across Process. Places Scales 1–465. https://doi.org/10.1017/CBO9781139235761

Boughton, W., 2006. Calibrations of a daily rainfall-runoff model with poor quality data. Environ. Model. Softw. 21, 1114–1128. https://doi.org/10.1016/J.ENVSOFT.2005.05.011

Burn, D.H., Boorman, D.B., 1993. Estimation of hydrological parameters at ungauged catchments. J. Hydrol. 143, 429–454. https://doi.org/10.1016/0022-1694(93)90203-L

Ceola, S., Arheimer, B., Baratti, E., Blöschl, G., Capell, R., Castellarin, A., Freer, J., Han, D., Hrachowitz, M., Hundecha, Y., Hutton, C., Lindström, G., Montanari, A., Nijzink, R., Parajka, J., Toth, E., Viglione, A., Wagener, T., 2015. Virtual laboratories: New opportunities for collaborative water science. Hydrol. Earth Syst. Sci. 19, 2101–2117. https://doi.org/10.5194/HESS-19-2101-2015

Chouaib, W., Alila, Y., Caldwell, P. V., 2018. Parameter transferability within homogeneous regions and comparisons with predictions from a priori parameters in the eastern United States. J. Hydrol. 560, 24–38. https://doi.org/10.1016/j.jhydrol.2018.03.018

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

Falkenmark, M., Chapman, T., 1989. Comparative Hydrology: An Ecological Approach to Land and Water Resources. Geography 27. https://doi.org/10.5860/CHOICE.27-3320

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)

He, Y., Bárdossy, A., Zehe, E., 2011. A review of regionalisation for continuous streamflow simulation. Hydrol. Earth Syst. Sci. 15, 3539–3553. https://doi.org/10.5194/hess-15-3539-2011

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

IRIMO, 2018. Temperature and precipitation dataset for Iran, Iran (in Persian).

Jahanshahi, A., Ghazanchaei, Z., Navari, M., Goharian, E., Patil, S.D., Zhang, Y., 2022. Dependence of rainfall-runoff model transferability on climate conditions in Iran. Hydrol. Sci. J. Accepted.

Jahanshahi, A., Melsen, L.A., Patil, S.D., Goharian, E., 2021. Comparing spatial and temporal scales of hydrologic model parameter transfer: A guide to four climates of Iran. J. Hydrol. 127099. https://doi.org/10.1016/J.JHYDROL.2021.127099

Kokkonen, T.S., Jakeman, A.J., Young, P.C., Koivusalo, H.J., 2003. Predicting daily flows in ungauged catchments: Model regionalization from catchment descriptors at the Coweeta

Hydrologic Laboratory, North Carolina. Hydrol. Process. 17, 2219–2238. https://doi.org/10.1002/hyp.1329

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

Neri, M., Parajka, J., Toth, E., 2020. Importance of the informative content in the study area when regionalising rainfall-runoff model parameters: The role of nested catchments and gauging station density. Hydrol. Earth Syst. Sci. 24, 5149–5171. https://doi.org/10.5194/HESS-24-5149-2020

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

Oudin, L., Kay, A., Andráassian, V., Perrin, C., 2010. Are seemingly physically similar catchments truly hydrologically similar? Water Resour. Res. 46. https://doi.org/10.1029/2009WR008887

Oudin, L., Perrin, C., Mathevet, T., Andréassian, V., Michel, C., 2006. Impact of biased and randomly corrupted inputs on the efficiency and the parameters of watershed models. J.

Hydrol. 320, 62–83. https://doi.org/10.1016/J.JHYDROL.2005.07.016

Parajka, J., Merz, R., Blöschl, G., 2007. Uncertainty and multiple objective calibration in regional water balance modelling: case study in 320 Austrian catchments. Hydrol. Process. 21, 435–446. https://doi.org/10.1002/HYP.6253

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., 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.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

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.

Petheram, C., Rustomji, P., Chiew, F.H.S., Vleeshouwer, J., 2012. Rainfall-runoff modelling in northern Australia: A guide to modelling strategies in the tropics. J. Hydrol. 462–463, 28–41. https://doi.org/10.1016/j.jhydrol.2011.12.046

Price, K., 2011. Effects of watershed topography, soils, land use and climate on baseflow hydrology in humid regions: a review. Prog. Physcial Geogr. 35(4)465-492 35, 465–492. https://doi.org/10.1177/0309133311402714

Sawicz, K.A., Kelleher, C., Wagener, T., Troch, P., Sivapalan, M., Carrillo, G., 2014. Characterizing hydrologic change through catchment classification. Hydrol. Earth Syst. Sci. 18, 273–285. https://doi.org/10.5194/HESS-18-273-2014

Singh, R., Archfield, S.A., Wagener, T., 2014. Identifying dominant controls on hydrologic parameter transfer from gauged to ungauged catchments - A comparative hydrology approach.

J. Hydrol. 517, 985–996. https://doi.org/10.1016/j.jhydrol.2014.06.030

Sivapalan, M., 2009. The secret to ‘doing better hydrological science’: change the question! Hydrol. Process. 23, 1391–1396. https://doi.org/10.1002/HYP.7242

Sivapalan, M., 2003. Process complexity at hillslope scale, process simplicity at the watershed scale: is there a connection? Hydrol. Process. 17, 1037–1041.

SKM, 2009. Regionalisation of hydrologic indices. Northern Australia sustainable yields. A Report Prepared by Sinclair Knight Merz for the CSIRO Northern Australia Sustainable Yields project, SKM. Melbourne.

Viglione, A., Parajka, J., 2019. TUWmodel: Lumped/SemiDistributed Hydrological Model for Education Purposes. R package version 1.1-0, available at: https: //CRAN.R-project.org/package=TUWmodel (last access: 26 October 2020).

Viney, N.R., Vaze, J., Chiew, F.H.S., Perraud, J., Post, D.A., Teng, J., 2009. Comparison of multi-model and multi-donor ensembles for regionalization of runoff generation using five lumped rainfall–runoff models. In: MODSIM 2009 International Congress on Modelling and Simulation. MSSANZ, Cairns, Australia, Cairns, Aust.

Wagener, T., Sivapalan, M., Troch, P., Woods, R., 2007. Catchment Classification and Hydrologic Similarity. Geogr. Compass 1, 901–931. https://doi.org/10.1111/j.1749-8198.2007.00039.x

Winter, T.C., 2001. The concept of hydrologic landscapes. J. Am. Water Resour. Assoc. 37, 335–349. https://doi.org/10.1111/J.1752-1688.2001.TB00973.X

Wolock, D.M., Winter, T.C., McMahon, G., 2004. Delineation and evaluation of hydrologic-landscape regions in the United States using geographic information system tools and multivariate statistical analyses. Environ. Manage. https://doi.org/10.1007/s00267-003-5077-9

Yang, W., Chen, H., Xu, C.Y., Huo, R., Chen, J., Guo, S., 2020. 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., 2020. 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

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