Dependence of rainfall-runoff model transferability on climate conditions in Iran

Jahanshahi, Afshin; Ghazanchaei, Zohreh; Navari, Mahdi; Goharian, Erfan; Patil, Sopan; Zhang, Yongqiang

Hydrological Sciences Journal

DOI:
10.1080/02626667.2022.2030867

Published: 12/03/2022

Peer reviewed version

Hydrological Sciences Journal

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/thsj20

Dependence of rainfall-runoff model transferability on climate conditions in Iran

Afshin Jahanshahi, Zohreh Ghazanchaei, Mahdi Navari, Erfan Goharian, Sopan D. Patil & Yongqiang Zhang

To cite this article: Afshin Jahanshahi, Zohreh Ghazanchaei, Mahdi Navari, Erfan Goharian, Sopan D. Patil & Yongqiang Zhang (2022): Dependence of rainfall-runoff model transferability on climate conditions in Iran, Hydrological Sciences Journal, DOI: 10.1080/02626667.2022.2030867

To link to this article: https://doi.org/10.1080/02626667.2022.2030867

Accepted author version posted online: 18 Jan 2022.

Publisher: Taylor & Francis & IAHS

Journal: *Hydrological Sciences Journal*

DOI: 10.1080/02626667.2022.2030867

Dependence of rainfall-runoff model transferability on climate conditions in Iran

Afshin Jahanshahi*[1], Zohreh Ghazanchaei[2], Mahdi Navari[3], Erfan Goharian[4], Sopan D. Patil[5], Yongqiang Zhang[6]

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- Zistab Consulting Engineers Co. Tehran, Iran.

3- University of Maryland Earth Systems Science Interdisciplinary Center and Hydrological Sciences Laboratory, Collage park, MD, USA.

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

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

6- Key Laboratory of Water Cycle and Related Land Surface Processes, Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, China.

Abstract

This study investigates the potential differences between spatial and temporal transferability of the HBV rainfall-runoff model parameters in 576 Iranian catchments. Our goal is to determine how parameter transferability is affected by (a) parameter transfer modes (temporal or spatial) and (b) different climate conditions. In the temporal mode, we decide in each catchment based on a benchmark baseline that accounts for the seasonality of flows. In the spatial mode, we examine physical similarity and spatial proximity methods. The main conclusions are that: (1) the HBV model struggles to beat the benchmark in lowland catchments, (2) under stationary conditions, rainfall-runoff models have greater temporal transferability than spatial transferability, (3) under non-stationary climate conditions, the number of catchments that perform best in terms of temporal transferability is smaller than in stationary climate conditions, and (4) the rainfall-runoff model has better transferability from drier to wetter conditions.

Keywords: Rainfall-runoff model, benchmark baseline, spatial and temporal transferability.

Introduction

Streamflow simulation in ungauged catchments remains a challenging task in hydrological science (Sivapalan *et al.*, 2003, Stoll & Weiler, 2010). Regionalization is a frequently used technique for estimating the amount of water resources in ungauged catchments. Regionalization involves the transfer of hydrological information (e.g. calibrated parameter sets) from gauged to ungauged or poorly gauged catchment(s) in order to estimate the streamflow time series (Razavi & Coulibaly, 2013). Over the last two decades, an increasing number of studies have applied different parameter transfer methods using hydrological models (e.g., Choubin *et al.*, 2019, Dakhlaoui *et al.*, 2017, Kokkonen *et al.*, 2003, Li *et al.*, 2014, Masih *et al.*, 2010, Perrin *et al.*, 2001, Reichl *et al.*, 2009, Vogel, 2005, Wagener & McIntyre, 2005, W. Yang *et al.*, 2020, Young, 2006).

Regionalization's main concern is with the operational application of hydrological models outside of calibration periods and at other catchments, where the parameter sets are subjected to actual examination (Patil & Stieglitz, 2015, Refsgaard & Knudsen, 1996). Parameter transfer, or regionalization, outside of calibration period can be in time (for simulating streamflow for periods

for which no observations are available) in space (for simulating in ungauged sites) or both (spatiotemporal).

The most frequently used approach in regionalization studies is the temporal transfer of model parameters (Patil & Stieglitz, 2015). Temporal transfer implicitly assumes that the calibrated parameters are temporally stable. Several recent studies, however, have demonstrated that calibrated model parameters are not temporally stable (e.g., Brigode *et al.*, 2013, Eregno *et al.*, 2013, Merz *et al.*, 2011, Yang *et al.*, 2018) and their values are determined by the calibration period's hydroclimatic conditions (Juston *et al.*, 2009, Razavi & Tolson, 2013).
A comprehensive review of the literature indicates that climate conditions are becoming non-stationary (IPCC 2014), and the reliability of the model transferability must be examined under climate variability. In recent decades, model transferability under stationary climate conditions has received considerable attention. However, accurate water resource forecasting under non-stationary climate conditions requires greater attention than ever before, as water-related problems such as drought and flooding are becoming more frequent as a result of global warming's effect (Li *et al.*, 2012). Non-stationary climate conditions also impose some constraints on the application of hydrological models, which may cause them to become less applicable in new conditions (KlemeŠ, 1986).

The Differential Split-Sample Test (DSST) is a popular method for assessing a model's transferability under contrasting conditions. DSST assumes that the observational data are divided into various calibration and validation periods based on their climatic differences (KlemeŠ, 1986). Numerous studies have revealed a significant decline in the model's transferability over time when using DSST (Fowler *et al.*, 2016, H. Li *et al.*, 2015, Wilby & Dessai, 2010).

Another widely used approach in regionalization studies is the spatial transfer of model parameters from gauge to ungauged catchments, often refer to as prediction in ungauged basins (PUB). Numerous studies have been conducted over the years to develop and compare methods for transferring model parameters between gauged and ungauged catchments (McIntyre *et al.*, 2005, Patil & Stieglitz, 2014, Sivapalan *et al.*, 2003, Yang *et al.*, 2018). Hrachowitz et al. (2013), Blöschl et al. (2013), and Parajka et al. (2013) provide a comprehensive overview of the achievements and discussions in PUB research during the PUB decade (2003-2012) initiative of the International Association of Hydrological Sciences (IAHS).

The transfer of calibrated model parameters between donor gauged and ungauged catchments cab based on Physical Similarity (PS) or Spatial Proximity (SP). They are the two most frequently used methods for spatial parameter transfer. The SP involves the transfer of calibrated parameters from neighboring gauged catchments to the ungauged catchment (Parajka *et al.*, 2005). The implicit hypothesis of the SP approach is that two adjacent catchments behave similarly in terms of hydrological response because their physical and climatic characteristics are likely to be similar (Chiew *et al.*, 2008, Petheram *et al.*, 2009). However, adjacent catchments may already have distinct characteristics and thus behave differently (Kennard *et al.*, 2010, Petheream & Bristow, 2008, Thornton *et al.*, 2007).

PS is the second method of parameter transfer (Kay *et al.*, 2007, Samaniego *et al.*, 2010). This method involves transferring calibrated parameters from the most physically similar catchment to the ungauged catchment (Bao *et al.*, 2012, Bárdossy, 2007, McIntyre *et al.*, 2005, Samuel *et al.*, 2011). However, determining which physical characteristics are necessary for successful parameter regionalization continues to be a challenge. The Euclidian distance is used to quantify the similarity of catchments (Kay et al., 2007). Catchment similarity is implemented in this paper using two of the most frequently used catchment descriptors (CDs), namely the aridity index and mean elevation.

Prediction of streamflow time series in ungauged catchments under contrasting climate conditions is increasingly becoming an important global challenge in hydrology. Iran is a case in a point of a country where regionalization may be critical due to: (a) its low gauge density and climatic diversity, and (b) a high likelihood of climate non-stationary due to impacts of climate change (Mansouri Daneshvar *et al.*, 2019). Interestingly, Iran also contains hydroclimatic regimes that are substantially different to those found in the other in France (e.g., Oudin *et al.*, 2008, 2010), Austria (e.g., Merz & Blöschl, 2004, Neri *et al.*, 2020, Parajka *et al.*, 2007), Australia (e.g., Li & Zhang, 2016, 2017, Zhang & Chiew, 2009), and the US (e.g., Patil and Stieglitz, 2014; Patil and Stieglitz, 2015), where a majority of PUB studies have been conducted in the past. Thus, the purpose of this study is to determine whether a hydrological model can be used to predict runoff time series under stationary and non-stationary conditions in Iran using the available dataset, covering an extensive range of climate types and conditions in data-scarce regions. To this end, the key objects of this study are to:

(1) Assess temporal transfer of model parameters under contrasting climate condition.

(2) Compare temporal and spatial parameter transfer under stationary and non-stationary climate conditions.

(3) Compare the physical similarity and spatial proximity methods for transferring model parameters from gauged to ungauged catchments.

2 Study area and dataset

2.1 Study area

The selection of the catchments throughout Iran is influenced by the aim of this study, i.e., to compare temporal and spatial transferability under stationary and non-stationary conditions. Therefore, the selected catchments must meet two requirements: (i) the availability of long-term hydro-climatic data for the same years period to make it possible to select contrasting climate conditions and (ii) the availability of high quality hydro-climate data with the minimum of data errors, and (iii) availability of physiographic characteristics for the study catchments. Thus, a sample of 576 Iranian catchments (Figure 1) is selected, which fulfilled these requirements. This catchment set covers a wide range of hydro climatic conditions and physiographic characteristics (Tables 1 and 2).

The dataset includes daily temperature, potential evapotranspiration, precipitation, and streamflow for the years 2000-2014. The mean annual precipitation (MAP) varies between 360 and 2000 mm/ year. Precipitation varies greatly across the country especially in the north, northwest, west, and central regions (from less than 400 to more than 2000 mm). The catchment area ranges from 65 to 12000 km^2. The elevation varies from -298 to 5595 m a. s. l. The average annual temperature (MAT) ranges from -3 to 39°C (IEM, 2018, IRIMO, 2018). Agriculture, rangeland, and forest are the three major land-use classes in the study catchments. From the northwest, northeast, and west to the central parts, the aridity increases (Appendix A presents aridity index and some physiographic characteristics at study catchments).

We classified the study catchments into three climate regions (arid, warm temperate, and snow) according to the Köppen-Geiger classification system (Figure 1). Figure 1a, shows that catchments in the snow class are located in Iran's western and northwestern parts. Warm temperate catchments can be found in Iran's northern, western, interior western, northeastern, and southwestern regions, while arid catchments can be found almost everywhere else. Severe precipitation and snow influence the hydrological regime in the snow and warm temperate classes, whereas rainfall influences the arid region. Using the k-means algorithm, the study catchments are also clustered into homogeneous areas (Figure 1b). The number of clusters (k = 3) is the same as the number of climate regions (Table 1).

2.2 Forcing dataset

2.2.1 Preparation of the precipitation and temperature data

Daily precipitation time series are collected for all catchments from point observations at gauge locations, but rainfall fields were estimated through the IDEW method (Appendix B presents the description of the IDEW method). Daily temperature time series are generated using a regression-based method using observations from the Iran Energy Ministry and Iran Meteorological Organization. Elevation is used as an explanatory factor. The Hargreaves method (Hargreaves *et al.*, 1985) is used to estimate reference evapotranspiration using maximum, minimum, and average temperatures. The time series of three model inputs was shown to be homogenous using the Standard Normal Homogeneity Test (SNHT) (Haimberger, 2007), with no breakpoints observed. The missing values in the data sets are estimated using the regression method using the values from neighboring gauges (elevation as an auxiliary variable). In general, a gauge's temperature data correlated well with corresponding data from neighboring gauges ($R^2 > 0.93$) that were used to fill in the missing records.

This correlation is $R^2 > 0.87$ in the case of precipitation data. On average, 7.8% and 10.4% of temperature and precipitation data for all 576 catchments, respectively, needed to be filled. The missing values in the discharge data set are also reconstructed using the drainage-area ratio technique proposed by Farmer and Vogel. (2013). On average, 4.1% of the discharge data for all 576 catchments had to be filled.

2.2.2 Hydro-climatic variability

The character of the entire period (2000-2014) is determined by hydro-climatic variables such as the mean monthly temperature (T), precipitation (P), PET (mm), and runoff (mm) across the various climate regions (arid, warm temperate, and snow regions).

Figure 2 shows the hydro-climatic attributes estimated using the period 2000-2014 for the 576 study catchments. These Figures demonstrate a high degree of inter-annual variability in precipitation followed by runoff, PET, and temperature in the study catchments from 2000 to 2014. The annual mean values for the hydro-climatic attributes are shown in Table 2 for three climate regions.

Figure 2a (right panel) shows that the MAP of the snow region varies between 461-616 mm/year; the MAP of the warm temperate region varies between 426-628 mm/year, and the MAP of the arid region varies between 380-591 mm/year. As we can see, MAT increased by an average of 0.55°C in the snow region, 0.32°C on average in the warm temperate region, and 0.09°C in the arid region. PET variability also correlates with temperature variability in three climate regions. Although runoff variability is smaller over water years, there are differences between climate regions (runoff in snow region varies between 41.7 and 120 mm/year, in the warm temperate region varies between 42.2 and 112.7 mm/year, and in the arid region varies between 40.2 and 82.6 mm/year).

2.3 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 three routines: snow, soil moisture, and flow response and routing (Ceola *et al.*, 2015, Parajka *et al.*, 2007, Viglione & Parajka, 2019). 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 (Appendix C). Finally, based on the sub-catchment areas, the different outputs from the elevation zones are averaged (Neri *et al.*, 2020).

The snow routine uses a simple degree-day notion to represent snow accumulation and melt, with a degree-day factor DDF and a melt temperature parameter T_M. A snow correction factor SCF is used to correct the catch deficit of precipitation gauges during snowfall. To distinguish between rainfall, snowfall, two threshold temperature parameters, T_R, and T_S are utilized. The soil moisture routine represents runoff generation and changes in the catchment's soil moisture state and includes three parameters: the maximum soil moisture storage FC, a parameter representing the soil moisture state above which evaporation occurs at its maximum rate, known as the limit for potential evaporation LP, and a parameter in the non-linear function relating runoff generation to the soil moisture state, known as the non-linearity parameter β. An upper and lower soil reservoir represents runoff routing on hillslopes. Excess rainfall enters the upper zone reservoir and exits through three routes: outflow from the reservoir using a fast storage coefficient K_1; percolation to the lower zone using a constant percolation rate C_{perc}; and, if a storage state l_{uz} threshold is exceeded, through an additional outlet using a very fast storage coefficient K_0. Water exits the lower zone based on a slow storage coefficient K_2. The outflow from both reservoirs is then routed using a triangle transfer function that reflects stream runoff routing, with the base of the transfer function, B_Q, estimated using the C_{ROUTE} and B_{MAX} parameters to scale the outflow (Neri *et al.*, 2020, Parajka *et al.*, 2007). Parajka et al. (2007) and Ceola et al. (2015), respectively, provide more details on the model structure and use in R.

3. Methodology

3.1. Differential and classical split-sample tests

To quantify the uncertainty associated with climate variability, this study employs a differential split-sample test (DSST). When a model is to be used to simulate streamflow under various climate conditions in gauged and ungauged catchments, DSST is (KlemeŠ, 1986, Ruelland *et al.*, 2015, Tramblay *et al.*, 2013, Vaze *et al.*, 2011, Westra *et al.*, 2014, W. Yang *et al.*, 2020, X. Yang *et al.*,

2018, Zheng *et al.*, 2018). This test is meaningful whenever a significant difference exists between the sampled wet/dry period climatic conditions and those founded in the available historical record. Thus, the sub-periods are composed of groups of hydroclimatically contrasted years in accordance with KlemeŠ (1986).

To create these sub-periods, the water years are divided into two categories based on the annual precipitation mean from 2000 to 2014 (Figure 2). Wet and dry years are defined as those with a MAP greater or less than the 30-year median. The differences in MAP and MAT range between -5.4% and +13.4% and from -13% to +3%, respectively, indicating a significant climatic contrast between the calibration and validation periods for DSST (Figure 2).

Additionally, this study employs a classical split sample test (CSST) (Yapo *et al.*, 1996) to determine which sub-periods have the best temporal transferability. This test is the most frequently used operation for evaluating model performance in stationary conditions. Our data period is divided into identical calibration and validation periods under this CSST. Therefore, we classified the water years into CSST (A and B) and DSST (C1, C2, D1, D2, E, and F) based on (i) climate-contrasted periods and (ii) the length of sub-periods (Figure 3). Numerous studies on model transferability have demonstrated that the optimal calibration period ranges between three and eight years when using a split sample test approach (Boughton, 2007, Li *et al.*, 2010, Yapo *et al.*, 1996).

3.2 Model calibration and evaluation

Adjusting the parameters of hydrological models is an essential part of hydrological simulations. The goodness-of-fit was improved by optimizing these parameter values until the difference between measured and simulated runoff was satisfactory (NSE $\geq$ 0.5) during model calibration.

Through the DEoptim package in R, the Differential Evolution optimization algorithm (DEoptim) (Storn & Price, 1997) was used to calibrate the model parameters (Ardia *et al.*, 2011). The algorithm is a member of the class of genetic algorithms that aim to maximize a specified objective function (Mitchell, 1998). DEoptim parameters were set to itermax = 400, population size (NP) = 400, trace = FALSE, crossover probability = 0.5, step-size = 0.8. The default strategy was used in the optimization procedure (DE/local-to-best/1/bin).

The DEoptim algorithm performed with six distinct maximum function calls set to 1000, 2000, 8000, 10,000, 15,000, and 20,000. Depending on the catchment, any number of function calls between 8000 and 10,000 produced the best results for the validation mode. There is only a slight variation in final performance between catchments calibrated for 8000 and 15,000 function calls, and the results are not significantly worse when the number of function calls equals 2000.

Model transferability was evaluated using Nash-Sutcliffe Efficiency NSE (Nash & Sutcliffe, 1970). The NSE criterion is a form of the normalized least-squares objective function. It places more emphasis on high flows. It is optimal when set to 1.

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

(1)

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

Preliminary analysis of catchment datasets was undertaken exhaustively and catchments with poor performance (for whatever reason) were excluded from the catchment dataset (997 preliminarily catchments reduced to 576 catchments). Out of 576 catchments in calibration, 11 catchments had lower calibration performance than other catchments (due to their proximity to karstic aquifers in western Iran). They are located in a semi-arid region with a mean NSE of 0.56 to 0.61 in calibration, but neighboring catchments with similar climatic conditions had a mean NSE of greater than 0.66. Under experiments, the calibration for these 11 catchments was utilized with random seeds (35 times). The standard deviation and lower and upper ranges of their parameters were examined for both training and testing data sets to determine that they were transferrable for regionalization. We repeated the process for the nearby catchments that performed better in the calibration. When the parameters of these two sets of catchments were analyzed, it was found that these 11 catchments are within a reliable range and can be used for regionalization.

3.3 Defining a benchmark for model performance

The NSE has an inherent benchmark in the form of the mean flow, giving NSE = 0 (Knoben *et al.*, 2020). When comparing model performance under different climate conditions, however, the definition of an appropriate benchmark is critical. This is especially important in model transferability studies conducted under a wide range of hydro-climate conditions (Parajka *et al.*, 2005). Here, we considered an NSE benchmark that was similar to Knoben et al. (2020).

To interpret model NSE values, we must first specify a benchmark score (Seibert *et al.*, 2018). The benchmark score is the lowest value we expect the model to achieve in each sub-period before deeming it suitable for the catchment under consideration.

The interannual mean value for every calendar day is a simple benchmark (Knoben *et al.*, 2020, Schaefli & Gupta, 2007, WMO, 1986). Thus, using data from the calibration periods, we calculate both the interannual median and mean flow per calendar day for each catchment. We then compare

the rainfall-runoff model's performance during the validation periods to the catchment's benchmark score. This benchmark NSE value represents the minimum level of accuracy we expect from the model before deeming acceptable for a given catchment.

We reported the difference between the model's NSE value and the benchmark value in each catchment, as well as the number of catchments where the model outperforms the benchmark. Figure 4 illustrates the detailed step-by-step implementation of all processes and model transferability.

4 Results

4.1 Model performance in both classical and differential split-sample tests

The CSST and DSST were used to determine calibration and validation periods in this study. These two tests are based on splitting periods into sub-periods, which show climatic conditions. We calibrated the model for 576 catchments in the CSST over two consecutive calibration periods (2000-2006 and 2007-2014), and then transferred the calibrated parameter sets to validation periods (2007-2014 and 2000-2006).

The boxplot of NSE values for the specified calibration, validation, PS, and SP over CSST and DSST are shown in Figure 5. The calibration and validation NSEs have changed over time as illustrated in this Figure. As expected, the median NSE values are highest during both CSST and DSST calibration periods. It is clear from this analysis that case A has superior temporal (validation mode) and spatial (PS and SP) transferability than case B over CSST. Except for E and F, temporal transferability outperformed spatial transferability in all DSST cases. As illustrated in Figure 5, the temporal transferability varies according to the climate contrasted period. As a result, this performance is still inferior in wetter to drier cases (B, C2, D2, E, and F) compared to drier to wetter cases (A, C1, and D1). This finding validated the model's reduced robustness in wet conditions when calibrated on dry conditions.

Overall, PS outperformed SP over both CSST and DSST. Comparing spatial and temporal transferability reveals that in the most extreme cases (i.e., E and F), PS and SP have lower median NSE values than validation. This superiority is more pronounced in the case of PS than in the case of SP. In these two cases, PS and SP are preferred over validation in catchments where contrasting climatic conditions or a temporal lag between the calibration and validation periods have resulted in poor temporal transferability. Generally, transferring calibrated parameters from a physically/geographically similar catchment to an ungauged catchment is preferable to transferring

calibrated parameters from a climate condition that is dissimilar to the ungauged catchment (see Figures 12g and 12 h in Section 4.6).

Additionally, Figure 5 shows the criteria of Moriasi et al. (2007). There are four performance categories: (i) very good (NSE > 0.75), (ii) good (0.65 <NSE ≤ 0.75), (ii) satisfactory (0.5 < NSE ≤ 0.65), and unsatisfactory (NSE ≤ 0.5). According to the criteria of Moriasi et al. (2007), the model exhibits satisfactory temporal and spatial transferability for cases A and B over CSST. For DSST, the model exhibits satisfactory temporal and spatial transferability in all cases except validation in C2, D1, and D2, where the model exhibits good temporal transferability.

4.2 Stability of model parameters over time

The Figures 6 and 7 illustrate a 1:1 comparison of 15 HBV model parameters for climate and homogeneous regions, respectively, over two calibration periods (2000-2006 and 2007-2014). Table 3 shows the r_1 correlation coefficients and median values of calibrated parameters for two periods. The degree-day factor (DDF) exhibits the greatest stability over time, as indicated by a correlation coefficient of 0.61. The weakest relationship is obtained for the storage coefficient for slow response, K_2, with a correlation coefficient of 0.31. This confirms that, as shown in section 4.1, not only model performance, but also parameter values, can vary significantly across calibration periods.

To determine the extent to which parameter variation between the two calibration periods can be attributed to differences in precipitation and temperature, we plotted the change in parameter value against ΔP and ΔT. (Figures 8 and 9). It was found that eight of the fifteen calibrated parameters (SCF, DDF, LP, K_1, K_2, FC, β, and Cperc) exhibit remarkable variation between the two calibration periods (Figure 5). At the country level, Figures 8 and 9 show that, for DP, parameter values show an increasing trend parallel with an increase in ΔP for all parameters except for SCF and Cperc (Figures 8a to 8h and 9a to 9h); for ΔT, the difference in all parameters shows an increasing trend parallel with an increase in ΔT with the exception of LP, FC, and Cperc.

4.3 Temporal and spatial transferability at regional scale

To increase the reliability of the results and investigate the preference of the model transferability to climate conditions, we implement parameter transfer at two climate-related and homogeneous regions (regional scale). Figure 10 shows the boxplots of calibration and model transferability for climate (Figure 10a) and homogenous (Figure 10b) regions over CSST and DSST.

For case A and B, the HBV model has superior temporal transferability over CSST in the arid class compared to snow and warm temperate. The arid class retains its superiority in temporal

transferability over DSST for C1 and D1. In general, PS has superior spatial transferability to SP when over CSST and DSST. Additionally, the Arid/c1 class/cluster exhibits superior spatial transferability to all other classes/clusters except A, C1, C2, and E. In general, the warm temperate/c3 shows less temporal and spatial transferability than other classes/clusters. The model is more transferable under D1 and D2, but less so under E and F, where there are temporal lag distances between the calibration and validation periods.

Additionally, the model's transferability is evaluated according to the criteria of Moriasi et al. (2007) criteria (dashed lines in Figure 10). This criterion further confirms that the arid and snow classes have superior temporal and spatial transferability to the warm temperate class (Figure 10a). Temporal transferability is greatest for D1 and D2, where the arid class has "very good" transferability. For PS, the arid and snow have "good" transferability for D1 and D2, whereas the snow class transfers satisfactory (Figure 10a).

E and F have the poorest temporal transferability of all the cases. The model exhibits satisfactory temporal transferability across all three climate regions, with arid and snow classes performing better than warm temperate (Figure 10a). These two cases' poor temporal transferability is explained by the contrasting climatic conditions and the temporal lag distance between the calibration and validation periods (see Figure 3), which means that weighting fractions are not uniform across all donor catchments. As a result of this finding, it is justified for E and F to have a worsened temporal transferability. This finding is consistent with the general findings by Patil and Stieglitz (2015) in the United States, where they found that the temporal parameter transfer method degrades as the temporal lag distance between the calibration and validation periods increases.

Overall, PS outperformed SP at the country level in the spatial mode. The snow and warm temperate classes continue to outperform the dry class in the spatial transfer mode. These two methods show better performance in D1, D2, C1, and F than in other cases. Figure 10b shows temporal and spatial transferability within homogeneous regions. c1 is more transferable in temporal mode than c2 and c3. Among all cases, c1 has the strongest temporal transferability under D1, while c2 and c3 have the weakest under E and F. In spatial mode, c1 retains its superiority over c2 and c3. Under A and B, C2 and c3 have the poorest spatial transferability. According to the criteria of Moriasi et al. (2007), the model is spatially transferrable for all clusters except c1, which is spatially transferrable under D1 and D2.

4.4 Defining benchmark NSE value

The spatial distribution of benchmark NSE values for the 576 catchments over the entire validation period (2000-2014) is shown in Figure 11. The performance of the benchmark values (i.e. the median or mean daily streamflow series) varies across space (Figure 11a). NSE values are lowest

(NSE ≤ 0) in the lowlands of western and southwestern Iran and are highest in the highlands throughout the country with values up to NSE = 0.815. There is no spatial coherence between climate regions in terms of benchmark values. Approximately 55% (n = 318) and 45% (n = 258) of these benchmarks are obtained using the mean and median calendar day streamflow, respectively (Figure 11b). These results set a threshold for minimum expected model temporal transferability by indicating the predictability of the streamflow regime is.

4.5 Model transferability over classical and differential split sample-tests

The NSE values of two cases (A and B) in validation over the CSST are shown in Figure 12 (left panel). Although there are expectations for the pattern, there is no strong spatial distribution. NSE values for validation range from 0.4 to 0.9. Geographically, temporal transferability is typically lowest in the west, central, and certain parts of the southeast (Figure 12a). These areas share to be arid and warm temperate (see Figure 1a). Figure 12b illustrates which catchments the model outperformed the benchmark NSE score.

Comparing model efficiency (validation mode) to pre-defined benchmark values in each catchment helps put the maximum model efficiency values into context (see Figure 12a/12c and Figure 12e/12f). The difference between model efficiency and benchmark score is smallest in mountainous catchments in the northwest, north, and western parts of the country and is generally greater in southwestern lowlands. For case A, in 130 catchments, the HBV model was unable to outperform the benchmark (validation ΔNSE is less than the benchmark score), the majority of which are characterized by a high fraction of snowfall occurring during cold seasons.

Figure 12a (case A) demonstrates that in the western and northwestern highlands, the number of catchments exceeding the benchmark, for whatever reason, is greater than in other parts. This partly explains the model includes a snow module that ensures the model can outperform the benchmark. Conversely, catchments in the interior, western, and northeastern parts of the arid and warm temperate regions are generally more difficult to model, as spatially variable rainfall events cause streamflow to vary in amplitude. Thus, the model must be applied with caution here. For case B (Figure 12c), the model performs worse than it does in the rest.

Here, climate conditions are not relatively similar in both cases. In case B, the calibration period's climate condition is wetter than the validation period's climate condition, whereas in case A, the climate condition is reversed (see Figure 3). In the case of B, the HBV model was not able to outperform the benchmark in 179 catchments. 95 out of 576 catchments are identical in these two cases (A and B) where the benchmark was not exceeded. Overall, our results indicate that the model is more transferable from drier to wetter conditions (case A) than from wetter to drier conditions (case B) over CSST.

The analysis of the model's performance change between calibration and validation reveals that performance change distributions are not similar across DSST. All wetter to drier cases (C2 and D2) exhibit less robust model performance than the drier to wetter cases (C1 and D1). The NSE values for climate contrasted periods (C1, C2, D1, and D2) in each catchment over the DSST are shown in Figure 13. As seen in these Figures, the number of catchments that exceed the benchmark in drier to wetter conditions is 80% (n = 461) and 92% (n = 534) for C1 and D1, respectively (Figures 13b and 13f). Conversely, in wetter to drier conditions (C2 and D2), the number of catchments that exceed the benchmark is decreases. This explains why the model is not transferable in at least 19% (n = 107) and 10% (n = 59) of the catchments for C2 and D2, respectively (Figures 13d and 13h).

To explore model capability more closely under contrasting climate conditions, we considered a four-year and an eight-year temporal lag distance between the calibration and validation periods in case E and case D, respectively. Between calibration and validation periods, here is the most contrasting climate condition in case F (see Figure 3). The results show that in these two cases the model has the least temporal transferability (see Figure 5b and Figures 13i to 13l). By comparing these two cases, we can deduce that the longer temporal lag distances (four and eight years, respectively) resulted in greater climatic contrasts between the calibration and validation periods, resulting in further degradation of temporal transferability.

Under all of these contrasting conditions, the model has lower temporal transferability in the northwestern and western parts of the country than in other regions (Figure 13, left panel). Case F has a lower transferability throughout Iran. The proportion of catchments exceeding the benchmark is greater in case E than in case F. (242 vs. 225). This explains why the model is inapplicable in these two extreme cases, at least in 58% (n = 334) and 61% (n = 351) of the study catchments, respectively (Figures 13j and 13l).

4.6 Comparing spatial and temporal parameter transfer modes

To determine the optimal mode of model transferability between CSST and DSST, we compare temporal and spatial transfer modes. Figure 14 shows the catchment locations where the temporal or spatial parameter transfer method performs best.

Following are three conclusions that can be drawn from this Figure: (i) no apparent coherence is noticeable regarding the catchments that prefer the temporal or spatial parameter transfer methods over the CSST and DSST, (ii) in general, PS outperformed SP in the majority of catchments. This

result demonstrates the significance of CDs (aridity index and mean elevation) over the geographical distance in the spatial mode. Thus, the superiority of PS over SP confirms the importance of catchment dynamics on CDs and (iii) temporal outperformed the spatial in a large number of catchments (ranging from n = 466 to n = 527) under stationary conditions (A, B, C1, C2, D1, and D2). Temporal still dominates spatial (both PS and SP) in non-stationary conditions (cases E and F), but in fewer catchments (n = 271 to n = 305). As a result, spatial transfer of model parameters from a physically (geographically) similar catchment to an ungauged catchment could be a viable option for only those catchments with limited model parameters temporal transferability. This finding contradicts the general conclusion by Patil and Stieglitz (2015) across the US, which found that when the temporal lag distance between calibration and validation periods is increased, the relative superiority of the temporal parameter transfer start behaving like spatial mode.

4.7 Controls on model transferability

The effect of two groups of static and dynamic CDs on the model transferability is investigated. The relationship between the model performance and (i) the climate variables' changes (mean annual temperature and mean annual precipitation) (ii) CDs (aridity index, elevation, and area) is presented below. Here, temporal transferability is assessed by calculating the difference between the calibration NSE (donor) and the NSE calculated in the same period but with calibrated parameters transferred from other periods (receiver).

Figure 15 shows the correlation matrix between differences in model performance between calibration and validation periods (DVAL) and differences in precipitation (DP), temperature (DT), catchment area, mean elevation, and aridity index at the local scale (the whole country). As seen in these Figures, DP has a positive correlation with DVAL in all eight cases. The main reason for that is that precipitation changes between calibration and validation periods have a significant impact on model transferability. The strongest correlation coefficients are found in cases E (r = 0.53) and F (r = 0.39). In other cases, the correlation is weak. Over CSST, case B has a stronger correlation than case A (r = 0.23 vs. r = 0.18). For all eight cases, there is a weak correlation between DT and DVAL, indicating that small temperature differences between calibration and validation periods (see Figure 3) have less impact on temporal transferability than DP (Appendix D presents the spatial distribution of DP and DT under all eight cases).

For B, C1, D1, and E, the aridity index has a positive correlation with DVAL; however, for A, C2, and F, it has a negative correlation. For D2, the correlation is zero. This correlation is positive for A, D1, and D2 in the elevation case, indicating that high-elevation catchments have better temporal

transferability. This correlation is positive for C2 and D2 in the catchment area case, implying that larger catchments have worse temporal transferability.

Figure 16 shows the relationship between the CDs and the difference in NSE values between calibration and spatial mode for physical similarity/spatial proximity (DPS/DSP). The difference between the calibration NSE (donor) and the NSE calculated in the same period but with calibrated parameters transferred under PS and SP is used to assess spatial transferability in this case (receiver). The aridity index and elevation have more control over spatial transferability in general.

5 Discussion

5.1 Assessment of spatial and temporal transferability

The model's performance was evaluated in this study by establishing benchmarks based on daily flow observations during the calibration period (Figure 11). This results in benchmark NSE values ranging from -0.23 and 0.815, with benchmark values in 55%of catchments being higher than what would be obtained using the mean annual flow. The benchmark score allows us to assess model performance by demonstrating the data's inter-annual variability and the efficiency values that can be easily obtained in a given catchment. Our findings reveal that when it comes to benchmark values, there is no clear spatial organization. The number of catchments that outperformed the benchmark is higher in highland catchments, as shown in Figures 11, 12, and 13 (NSE values and benchmark score). This explains why the dense networks of measuring gauges ensure that the model will outperform the benchmark.

Two regionalization methods are used to assess spatial transferability. Physical similarity outperformed spatial proximity across Iran. This finding confirms that catchment descriptors (aridity index and mean elevation) have a greater impact on transferring parameters from gauged to ungauged catchments than the geographical distance between their centroids. This discrepancy in performance could be due to a lack of streamflow and meteorological network in some parts of the country, making it difficult to choose the best parameter transfer strategy between climate regions. Given the conclusion by Oudin et al. (2008), it is impossible to determine the most robust regionalization method when the streaming network density is less than 60 gauges per 100,000 km^2.

For a more detailed analysis, we examined the temporal and spatial transferability of different climate/homogeneous regions of Iranian catchments. The regionalization performance is shown in Figure 10 for all climate (arid, warm temperate, and snow) and homogenous (c1, c2, and c3) regions. The overall performance of regionalization is inconsistent with the regionalization studies in Parajka et al. (2013). Almost all regionalization studies have concluded that the distance-based regionalization methods perform better in humid and cold catchments (in this case, snow class and c3 cluster) than in arid catchments (in this case, arid/warm temperate classes and c1/c2 cluster). Additionally, they concluded that PS and SP perform similarly; however, our results demonstrated that the PS is superior to SP.

5.2 Temporal model transferability under non-stationary climate conditions

This study demonstrates that the difference in climate conditions between calibration and validation periods gradually affects the model's transferability when applied to Iranian catchments under contrasted-climatic conditions. The HBV rainfall-runoff model is found to be transferable to colder and/or wetter conditions. However, its transferability deteriorated when climate conditions involved an increase of more than 14% in MAP (see case F in Figure 3) and an 8-year temporal lag distance between calibration and validation. A similar conclusion was found by Dakhlaoui et al. (2017) in northern Tunisia and Coron et al. (2012) in southwest Australia, Sleziak et al. (2016) in Austria, and Oudin et al. (2006) in the US, who concluded that a significantly increasing trend in precipitation resulted in a decreasing trend in model transferability. Our findings suggest that, hydrological modelers should prefer similar calibration and verification periods in simulation studies. This would significantly maintain model's robustness in non-stationary conditions. It is also confirmed by the benchmark values that the model calibrated in wet conditions has no temporal transferability to dry conditions (Figures 13).

The validation results indicate that the number of mountainous catchments (highlands) exceeding the benchmark is greater than the number of low-elevation catchments (lowlands). This finding confirms that both the term ‘acceptable model’ and the term ‘superior performance’ are ascribed to climatic conditions (Figures 12 and 13).

Model performance in terms of temporal transferability is highly dependent on our definition of “acceptable” and “not acceptable” model performance. Without an explicit statement about our expected benchmark values, we might have concluded that our model performs worse in lowlands, although expectations to the pattern exist (see negative values in Figure 11). In fact, this would have been following literature that states that it is harder to obtain high-efficiency values in arid catchments than it is to obtain such values in humid locations (e.g. Fowler *et al.*, 2018, Knoben *et al.*, 2020, Krysanova *et al.*, 2017, Melsen *et al.*, 2018). There is no robust spatial coherence in arid

or wet catchments where our model performs poorly or strongly, using our specified benchmark. Despite the difficulties encountered by hydrological modelers in arid regions (Pilgrim *et al.*, 1988), our model outperforms the benchmark in both humid and arid regions, but not equally.

5.3 Effect of climate variables and catchment descriptors on model transferability

This study demonstrates that two distinct types of catchments descriptors, dynamic and static, have varying effects on temporal and spatial transferability. The relationship between model transferability and these catchments descriptors is more complicated and depends on the region-specific.

For example, in terms of temporal transferability throughout Iran, precipitation changes between calibration and validation are more significant than temperature changes (Figure 15). Dakhlaoui et al. (2017) confirmed this finding in northern Tunisia. They demonstrated that the tested models are incapable of temporal transfer when climate conditions involve an increase of more than +1.75 °C in annual mean temperatures and a decrease of more than 25% in annual precipitation.

In France (Oudin *et al.*, 2008) and Australia (Z. Zhang *et al.*, 2008), performance decreases with increasing mean elevation, whereas performance increases with increasing mean elevation in Austria (Parajka *et al.* 2005). Our research does not corroborate this finding. Here, the pattern of air temperatures is similar to that of the aridity index, which contradicts the study in Austria (Parajka *et al.*, 2005) but is consistent with Oudin *et al.* (2008) in France. Our findings confirm that both climate conditions and regionalization methods have an effect on the model's transferability. Future work should focus on the effect of the land-use changes/soil characteristics on model robustness under climatic variability, as this will help to explain some of the human impact on model robustness.

5.4 Transferability of models at the local and regional levels

Model transferability at local and regional scales indicates that arid and snow climate regions, as well as the c1 and c2 clusters, have higher temporal and spatial transferability than the entire country (local or all study catchments combined), whereas the warm temperate climate region and the c3 cluster have lower transferability than the entire country (compare Figures 5 and 10). Thus, the model's superior at the regional scale confirmed a degree of homogeneity in both the state variables (snow, soil moisture, etc.) and the model's response to the same meteorological forcing. This conclusion is also consistent with the conclusion by Lachance-Cloutier *et al.* (2017) in Quebec, Canada. According to Parajka *et al.* (2013) regionalization studies involving a large number of catchments produced better results than studies involving a small number of catchments throughout the world. Our study does not corroborate this finding.

Due to the fact that the median nearest neighbor distance at the local scale is different from the median nearest neighbor distance at the regional scale, it can be concluded that the geographical distance between the donor and target catchments has a significant effect on the results. That is why the SP results for these two scales differ (the regional is somewhat superior).

6 Conclusions

This study's main aim was to determine the effect of various combination of climate conditions, spatial and temporal transferability on the performance of streamflow simulations in ungauged catchments. We calibrated HBV conceptual rainfall-runoff model in 576 catchments covering three climate regions according to the Köppen-Geiger classification, with the Nash-Sutcliff Efficiency (NSE) as the objective function. To determine the model's efficiency for each catchment, a benchmark was defined based on the median or mean calendar day streamflow. The robustness of the HBV model in temporal mode was investigated using a differential split-sample test (DSST) and classical split sample-test (CSST) based on the climate classification of the observation period.

Temperature and precipitation, were used as climate variables to classify the study period into climate-contrasted periods. The results indicated that the contrasting climatic conditions between calibration and validation periods gradually reduced temporal transferability. While our study model was transferable to colder and/or wetter conditions, its efficiency degradation was greater when precipitation decreased than when precipitation increased. The transferability studies should take into account the similarity in climatic conditions between the calibration and validation periods in order to reduce the simulation uncertainty. In highlands with contrasting climate conditions, the benchmark baseline proved easiest to beat. The model can outperform the benchmark in both arid and wet catchments, but the benchmark must be carefully chosen.

Under stationary climate conditions, we investigated the performance of the physical similarity and spatial proximity regionalization methods in spatial mode. We found that physical similarity outperformed spatial proximity in a larger number of catchments.

The HBV model was more transferable in the temporal than spatial mode; however, there does not appear to be any coherence in transferability over two CSST and DSST. In cases where the difference between calibration and validation climate conditions is greatest (most extreme cases), the proportion of catchments that exceed the benchmark score is lower than in other non-stationary conditions. Temporal predominates over spatial in these most extreme cases, but in a smaller number of catchments than the stationary condition. Comparing model transferability across climate regions revealed that arid climate regions have greater transferability than snow and warm temperate climate regions.

Comparing the effects of static (mean elevation and catchment area) and dynamic (aridity index and difference in precipitation/temperature between calibration and validation periods) catchment descriptors on model transferability revealed that dynamics had a greater effect. Additionally, the results of this study confirmed that daily data can be used as a reliable scale for the PUB paradox of transferability in an arid and semi-arid country like Iran.

The insights acquired in this study provide a useful reference for adapting calibration and validation periods in order to achieve the most appropriate mode of model parameter transfer. We recommend incorporating additional approaches and hydrological models to demonstrate the effect of model selection on model transferability.

Acknowledgements

The authors are grateful to Iran Energy Ministry, Iran Meteorological Organization, and Water Resources Investigation consulting engineers for their cooperation and provision of required data. We would also like to thank Dr. Jan Magnusson, whose detailed comments led to significant improvements to the original manuscript.

Conflicts of interest

The authors declare no conflict of interest.

Data availability

https://www.dropbox.com/s/rjog04at8gd9ute/Data.zip?dl=0

References

Ardia, D., Boudt, K., Carl, P., Mullen, K. M. & Peterson, B. G. (2011) Differential evolution with deoptim. *R J.* **3**(1), 27–34. doi:10.32614/rj-2011-005

Bao, Z., Zhang, J., Liu, J., Fu, G., Wang, G., He, R., Yan, X., et al. (2012) Comparison of regionalization approaches based on regression and similarity for predictions in ungauged catchments under multiple hydro-climatic conditions. *J. Hydrol.* **466–467**, 37–46. Elsevier B.V. doi:10.1016/j.jhydrol.2012.07.048

Bárdossy, A. (2007) Calibration of hydrological model parameters for ungauged catchments. *Hydrol. Earth Syst. Sci.* **11**(2), 703–710. doi:10.5194/hess-11-703-2007

Boughton, W. C. (2007) Effect of data length on rainfall-runoff modelling. *Environ. Model. Softw.* **22**(3), 406–413. doi:10.1016/J.ENVSOFT.2006.01.001

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

Ceola, S., Arheimer, B., Baratti, E., Blöschl, G., Capell, R., Castellarin, A., Freer, J., et al. (2015) Virtual laboratories: New opportunities for collaborative water science. *Hydrol. Earth Syst. Sci.* **19**(4), 2101–2117. Copernicus GmbH. doi:10.5194/HESS-19-2101-2015

Chiew, F. H. S., Vaze, J., Viney, N., Jordan, P., Perraud, J., Zhang, L., Teng, J., et al. (2008) Rainfall-runoff modelling across the Murray-Darling Basin. Rainfall-runoff modelling across the Murray-Darling Basin. A Report to the Australian Government from the CSIRO Murray-Darling Basin Sustainable Yields Project. *CSIRO*. Retrieved from https://publications.csiro.au/rpr/pub?list=SEA&pid=procite:8cb03ed2-4bbd-4801-b570-34a27fd70495

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

Dakhlaoui, H., Ruelland, D., Tramblay, Y. & Bargaoui, Z. (2017) Evaluating the robustness of conceptual rainfall-runoff models under climate variability in northern Tunisia. *J. Hydrol.* **550**, 201–217. doi:10.1016/j.jhydrol.2017.04.032

Eregno, F. E., Xu, C. Y. & Kitterød, N. O. (2013) Modeling hydrological impacts of climate change in different climatic zones. *Int. J. Clim. Chang. Strateg. Manag.* **5**(3), 344–365. Emerald Group Publishing Limited. doi:10.1108/IJCCSM-04-2012-0024

Fowler, K. J. A., Peel, M. C., Western, A. W., Zhang, L. & Peterson, T. J. (2016) Simulating runoff under changing climatic conditions: Revisiting an apparent deficiency of conceptual rainfall-runoff models. *Water Resour. Res.* **52**(3), 1820–1846. Blackwell Publishing Ltd. doi:10.1002/2015WR018068

Fowler, K., Peel, M., Western, A. & Zhang, L. (2018) Improved Rainfall-Runoff Calibration for Drying Climate: Choice of Objective Function. *Water Resour. Res.* **54**(5), 3392–3408. Blackwell Publishing Ltd. doi:10.1029/2017WR022466

Haimberger, L. (2007) Homogenization of radiosonde temperature time series using innovation statistics. *J. Clim.* **20**(7), 1377–1403. American Meteorological Society. doi:10.1175/JCLI4050.1

Hargreaves, G. L., Hargreaves, G. H. & Riley, J. P. (1985) Irrigation Water Requirements for Senegal River Basin. *J. Irrig. Drain. Eng.* **111**(3), 265–275. American Society of Civil Engineers (ASCE). doi:10.1061/(asce)0733-9437(1985)111:3(265)

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

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

Juston, J., Seibert, J. & Johansson, P.-O. (2009) Temporal sampling strategies and uncertainty in calibrating a conceptual hydrological model for a small boreal catchment. *Hydrol. Process.* **23**(21), 3093–3109. John Wiley & Sons, Ltd. doi:10.1002/hyp.7421

Kay, A. L., Jones, D. A., Crooks, S. M., Kjeldsen, T. R. & Fung, C. F. (2007) An investigation of site-similarity approaches to generalisation of a rainfall-runoff model. *Hydrol. Earth Syst. Sci.* **11**(1), 500–515. Copernicus GmbH. doi:10.5194/hess-11-500-2007

Kennard, M. J., Pusey, B. J., Olden, J. D., MacKay, S. J., Stein, J. L. & Marsh, N. (2010) Classification of natural flow regimes in Australia to support environmental flow management. *Freshw. Biol.* **55**(1), 171–193. John Wiley & Sons, Ltd. doi:10.1111/j.1365-2427.2009.02307.x

KlemeŠ, V. (1986) Operational testing of hydrological simulation models. *Hydrol. Sci. J.* **31**(1), 13–24. Taylor & Francis Group . doi:10.1080/02626668609491024

Knoben, W. J. M., Freer, J. E., Peel, M. C., Fowler, K. J. A. & Woods, R. A. (2020) A Brief Analysis of Conceptual Model Structure Uncertainty Using 36 Models and 559 Catchments. *Water Resour. Res.* **56**(9). doi:10.1029/2019WR025975

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**(11), 2219–2238. doi:10.1002/hyp.1329

Krysanova, V., Vetter, T., Eisner, S., Huang, S., Pechlivanidis, I., Strauch, M., Gelfan, A., et al. (2017) Intercomparison of regional-scale hydrological models and climate change impacts projected for 12 large river basins worldwide - A synthesis. *Environ. Res. Lett.* **12**(10), 105002. Institute of Physics Publishing. doi:10.1088/1748-9326/aa8359

Li, C. Z., Zhang, L., Wang, H., Zhang, Y. Q., Yu, F. L. & Yan, D. H. (2012) The transferability of hydrological models under nonstationary climatic conditions. *Hydrol. Earth Syst. Sci.* **16**(4), 1239–1254. doi:10.5194/hess-16-1239-2012

Li, Chuan Zhe, Wang, H., Liu, J., Yan, D. H., Yu, F. L. & Zhang, L. (2010) Effect of calibration data series length on performance and optimal parameters of hydrological model. *Water Sci. Eng.* **3**(4), 378–393. Elsevier. doi:10.3882/J.ISSN.1674-2370.2010.04.002

Li, F., Zhang, Y., Xu, Z., Liu, C., Zhou, Y. & Liu, W. (2014) Runoff predictions in ungauged catchments in southeast Tibetan Plateau. *J. Hydrol.* **511**, 28–38. Elsevier B.V. doi:10.1016/j.jhydrol.2014.01.014

Li, Hong, Beldring, S. & Xu, C.-Y. (2015) Stability of model performance and parameter values on

two catchments facing changes in climatic conditions. *Hydrol. Sci. J.* **60**(7–8), 1317–1330. Taylor and Francis Ltd. doi:10.1080/02626667.2014.978333

Li, Hongxia & Zhang, Y. (2016) Regionalising rainfall–runoff modelling for predicting daily runoff in continental Australia. *Hydrol. Earth Syst. Sci. Discuss.* (September), 1–24. doi:10.5194/hess-2016-464

Li, Hongxia & Zhang, Y. (2017) Regionalising rainfall-runoff modelling for predicting daily runoff: Comparing gridded spatial proximity and gridded integrated similarity approaches against their lumped counterparts. *J. Hydrol.* **550**, 279–293. Elsevier B.V. doi:10.1016/j.jhydrol.2017.05.015

Mansouri Daneshvar, M. R., Ebrahimi, M. & Nejadsoleymani, H. (2019) An overview of climate change in Iran: facts and statistics. *Environ. Syst. Res. 2019 81* **8**(1), 1–10. SpringerOpen. doi:10.1186/S40068-019-0135-3

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**(1–2), 188–201. Elsevier B.V. doi: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**(12), 1–14. doi:10.1029/2005WR004289

Melsen, L. A., Addor, N., Mizukami, N., Newman, A. J., Torfs, P. J. J. F., Clark, M. P., Uijlenhoet, R., et al. (2018) Mapping (dis)agreement in hydrologic projections. *Hydrol. Earth Syst. Sci.* **22**(3), 1775–1791. Copernicus GmbH. doi:10.5194/hess-22-1775-2018

Merz, R. & Blöschl, G. (2004) Regionalisation of catchment model parameters. *J. Hydrol.* **287**(1–4), 95–123. Elsevier. doi:10.1016/j.jhydrol.2003.09.028

Merz, R., Parajka, J. & Blöschl, G. (2011) Time stability of catchment model parameters: Implications for climate impact analyses. *Water Resour. Res.* **47**(2), 1–17. doi:10.1029/2010WR009505

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

Nash, J. E. & Sutcliffe, J. V. (1970) River flow forecasting through conceptual models part I - A discussion of principles. *J. Hydrol.* **10**(3), 282–290. Elsevier. doi:10.1016/0022-1694(70)90255-6

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**(11), 5149–5171. Copernicus GmbH.

doi:10.5194/HESS-24-5149-2020

Oudin, L., Andréassian, V., Perrin, C., Michel, C. & Moine, N. Le. (2008) Spatial proximity, physical similarity, regression and ungaged catchments: A comparison of regionalization approaches based on 913 French catchments. *Water Resour. Res.* **44**(3), 1–15. doi:10.1029/2007WR006240

Oudin, L., Kay, A., Andráassian, V. & Perrin, C. (2010) Are seemingly physically similar catchments truly hydrologically similar? *Water Resour. Res.* **46**(11). doi:10.1029/2009WR008887

Parajka, J., Merz, R. & Blöschl, G. (2005) A comparison of regionalisation methods for catchment model parameters. *Hydrol. Earth Syst. Sci.* **9**(3), 157–171. doi:10.5194/hess-9-157-2005

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**(4), 435–446. John Wiley & Sons, Ltd. doi:10.1002/hyp.6253

Patil, S. D. & Stieglitz, M. (2015) Comparing spatial and temporal transferability of hydrological model parameters. *J. Hydrol.* **525**, 409–417. Elsevier B.V. doi:10.1016/j.jhydrol.2015.04.003

Patil, S. & Stieglitz, M. (2014) Modelling daily streamflow at ungauged catchments: what information is necessary? *Hydrol. Process.* **28**(3), 1159–1169. John Wiley & Sons, Ltd. doi:10.1002/hyp.9660

Perrin, C., Michel, C. & Andréassian, V. (2001) Does a large number of parameters enhance model performance? Comparative assessment of common catchment model structures on 429 catchments. *J. Hydrol.* **242**(3–4), 275–301. Elsevier. doi:10.1016/S0022-1694(00)00393-0

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

Pethream, C. & Bristow, K. . (2008) Towards an understanding of the hydrological factors, constraints and opportunities for irrigation in northern Australia: A review. *CSIRO L. Water Sci. Rep.* Retrieved March 31, 2021, from https://publications.csiro.au/rpr/pub?pid=procite:99530409-31bc-4570-97a6-83d715506c56

Pilgrim, D. H., Chapman, T. G. & Doran, D. G. (1988) Problèmes de la mise au point de modèles pluie-écoulement dans les régions arides et semi-arides. *Hydrol. Sci. J.* **33**(4), 379–400. Taylor & Francis Group . doi:10.1080/02626668809491261

Razavi, S. & Tolson, B. A. (2013) An efficient framework for hydrologic model calibration on long data periods. *Water Resour. Res.* **49**(12), 8418–8431. John Wiley & Sons, Ltd. doi:10.1002/2012WR013442

Razavi, T. & Coulibaly, P. (2013) Streamflow Prediction in Ungauged Basins: Review of

Regionalization Methods. *J. Hydrol. Eng.* **18**(8), 958–975. doi:10.1061/(asce)he.1943-5584.0000690

Refsgaard, J. C. & Knudsen, J. (1996) Operational Validation and Intercomparison of Different Types of Hydrological Models. *Water Resour. Res.* **32**(7), 2189–2202. John Wiley & Sons, Ltd. doi:10.1029/96WR00896

Reichl, J. P. C., Western, A. W., McIntyre, N. R. & Chiew, F. H. S. (2009) Optimization of a similarity measure for estimating ungauged streamflow. *Water Resour. Res.* **45**(10). John Wiley & Sons, Ltd. doi:10.1029/2008WR007248

Ruelland, D., Hublart, P. & Tramblay, Y. (2015) Assessing uncertainties in climate change impacts on runoff in Western Mediterranean basins. *IAHS-AISH Proc. Reports* **371**, 75–81. doi:10.5194/piahs-371-75-2015

Samaniego, L., Bárdossy, A. & Kumar, R. (2010) Streamflow prediction in ungauged catchments using copula-based dissimilarity measures. *Water Resour. Res.* **46**(2). John Wiley & Sons, Ltd. doi:10.1029/2008WR007695

Samuel, J., Coulibaly, P. & Metcalfe, R. A. (2011) Estimation of Continuous Streamflow in Ontario Ungauged Basins: Comparison of Regionalization Methods. *J. Hydrol. Eng.* **16**(5), 447–459. American Society of Civil Engineers (ASCE). doi:10.1061/(asce)he.1943-5584.0000338

Schaefli, B. & Gupta, H. V. (2007) Do Nash values have value? *Hydrol. Process.* **21**(15), 2075–2080. John Wiley & Sons, Ltd. doi:10.1002/hyp.6825

Seibert, J., Vis, M. J. P., Lewis, E. & Meerveld, H. J. van. (2018) Upper and lower benchmarks in hydrological modelling. *Hydrol. Process.* **32**(8), 1120–1125. John Wiley and Sons Ltd. doi:10.1002/hyp.11476

Sivapalan, M., Takeuchi, K., Franks, S. W., Gupta, V. K., Karambiri, H., Lakshmi, V., Liang, X., et al. (2003) IAHS Decade on Predictions in Ungauged Basins (PUB), 2003-2012: Shaping an exciting future for the hydrological sciences. *Hydrol. Sci. J.* **48**(6), 857–880. Taylor & Francis Group . doi:10.1623/hysj.48.6.857.51421

Stoll, S. & Weiler, M. (2010) Explicit simulations of stream networks to guide hydrological modelling in ungauged basins. *Hydrol. Earth Syst. Sci.* **14**(8), 1435–1448. doi:10.5194/hess-14-1435-2010

Storn, R. & Price, K. (1997) Differential Evolution - A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces. *J. Glob. Optim.* **11**(4), 341–359. Springer Netherlands. doi:10.1023/A:1008202821328

Thornton, C. M., Cowie, B. A., Freebairn, D. M. & Playford, C. L. (2007) The Brigalow Catchment Study: II. Clearing brigalow (Acacia harpophylla) for cropping or pasture increases runoff. *Soil Res.* **45**(7), 496. CSIRO PUBLISHING. doi:10.1071/SR07064

Tramblay, Y., Ruelland, D., Somot, S., Bouaicha, R. & Servat, E. (2013) High-resolution Med-CORDEX regional climate model simulations for hydrological impact studies: A first evaluation of the ALADIN-Climate model in Morocco. *Hydrol. Earth Syst. Sci.* **17**(10), 3721–3739. doi:10.5194/hess-17-3721-2013

Vaze, J., Davidson, A., Teng, J. & Podger, G. (2011) Impact of climate change on water availability in the Macquarie-Castlereagh River Basin in Australia. *Hydrol. Process.* **25**(16), 2597–2612. John Wiley & Sons, Ltd. doi:10.1002/hyp.8030

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).

Vogel, R. M. (2005) Regional calibration of watershed models. In: *Watershed Models*, 47–71. CRC Press. doi:10.1201/9781420037432.ch3

Wagener, T. & McIntyre, N. (2005) Identification of rainfall-runoff models for operational applications. *Hydrol. Sci. J.* **50**(5), 735–751. IAHS Press . doi:10.1623/hysj.2005.50.5.735

Westra, S., Thyer, M., Leonard, M., Kavetski, D. & Lambert, M. (2014) A strategy for diagnosing and interpreting hydrological model nonstationarity. *Water Resour. Res.* **50**(6), 5090–5113. Blackwell Publishing Ltd. doi:10.1002/2013WR014719

Wilby, R. L. & Dessai, S. (2010) Robust adaptation to climate change. *Weather* **65**(7), 180–185. John Wiley & Sons, Ltd. doi:10.1002/wea.543

WMO, W. M. O. (1986) *Intercomparison of models of snowmelt runoff*. WMO. Geneva: WMO.

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. Elsevier B.V. doi:10.1016/j.jhydrol.2020.125276

Yang, X., Magnusson, J., Rizzi, J. & Xu, C. Y. (2018) Runoff prediction in ungauged catchments in Norway: Comparison of regionalization approaches. *Hydrol. Res.* **49**(2), 487–505. doi:10.2166/nh.2017.071

Yapo, P. O., Gupta, H. V. & Sorooshian, S. (1996) Automatic calibration of conceptual rainfall-runoff models: Sensitivity to calibration data. *J. Hydrol.* **181**(1–4), 23–48. Elsevier. doi:10.1016/0022-1694(95)02918-4

Young, A. R. (2006) Stream flow simulation within UK ungauged catchments using a daily rainfall-runoff model. *J. Hydrol.* **320**(1–2), 155–172. doi:10.1016/j.jhydrol.2005.07.017

Zhang, Y. & Chiew, F. H. S. (2009) Relative merits of different methods for runoff predictions in ungauged catchments. *Water Resour. Res.* **45**(7). doi:10.1029/2008WR007504

Zhang, Z., Wagener, T., Reed, P. & Bhushan, R. (2008) Reducing uncertainty in predictions in ungauged basins by combining hydrologic indices regionalization and multiobjective

optimization. *Water Resour. Res.* **44**(12), 1–13. doi:10.1029/2008wr006833

Zheng, F., Maier, H. R., Wu, W., Dandy, G. C., Gupta, H. V. & Zhang, T. (2018) On Lack of Robustness in Hydrological Model Development Due to Absence of Guidelines for Selecting Calibration and Evaluation Data: Demonstration for Data-Driven Models. *Water Resour. Res.* **54**(2), 1013–1030. Blackwell Publishing Ltd. doi:10.1002/2017WR021470

Appendix A

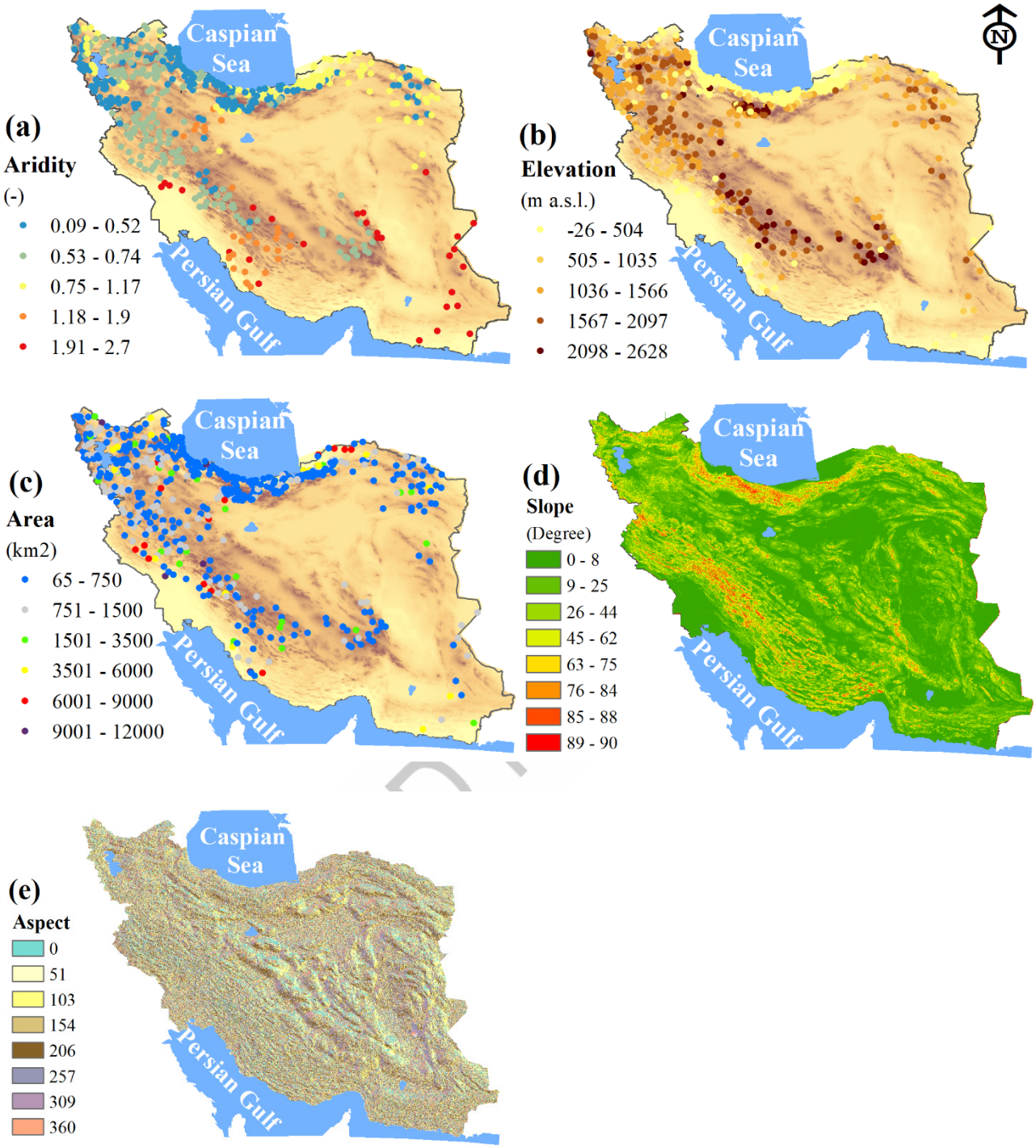


Figure A1. Spatial distribution of aridity, elevation, area, slope and aspect at 576 study catchments.

Appendix B

Preparation of precipitation 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 IDW and Elevation (IDEW) method. The IDEW is an interpolation technique and offers the possibility of defining elevation and distance weighting, making it more suitable for mountainous regions of Iran. This technique was shown to be more suitable for mountainous catchments in the Karkheh River Basin and southwestern Iran (Masih et al., 2011, 2010; Modallakdoust et al., 2008). The equation for this method is as follows: $\hat{P}_k = W_D \sum_{i=1}^{N} \frac{1}{D} w\,(d)_i p_i + W_Z \sum_{i=1}^{N} \frac{1}{Z} w\,(Z)_i p_i$

(A1)

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

$w(d) = 1/d^a \quad for \quad d > 0$

(A2)

$$f(x) = \begin{cases} 1/z_{min}^b & for \quad z \le z_{min} \\ 1/z^b & for \quad z_{min} < z < z_{max} \\ 0 & for \quad z \ge z_{max} \end{cases}$$

(A3)

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

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

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

Appendix C

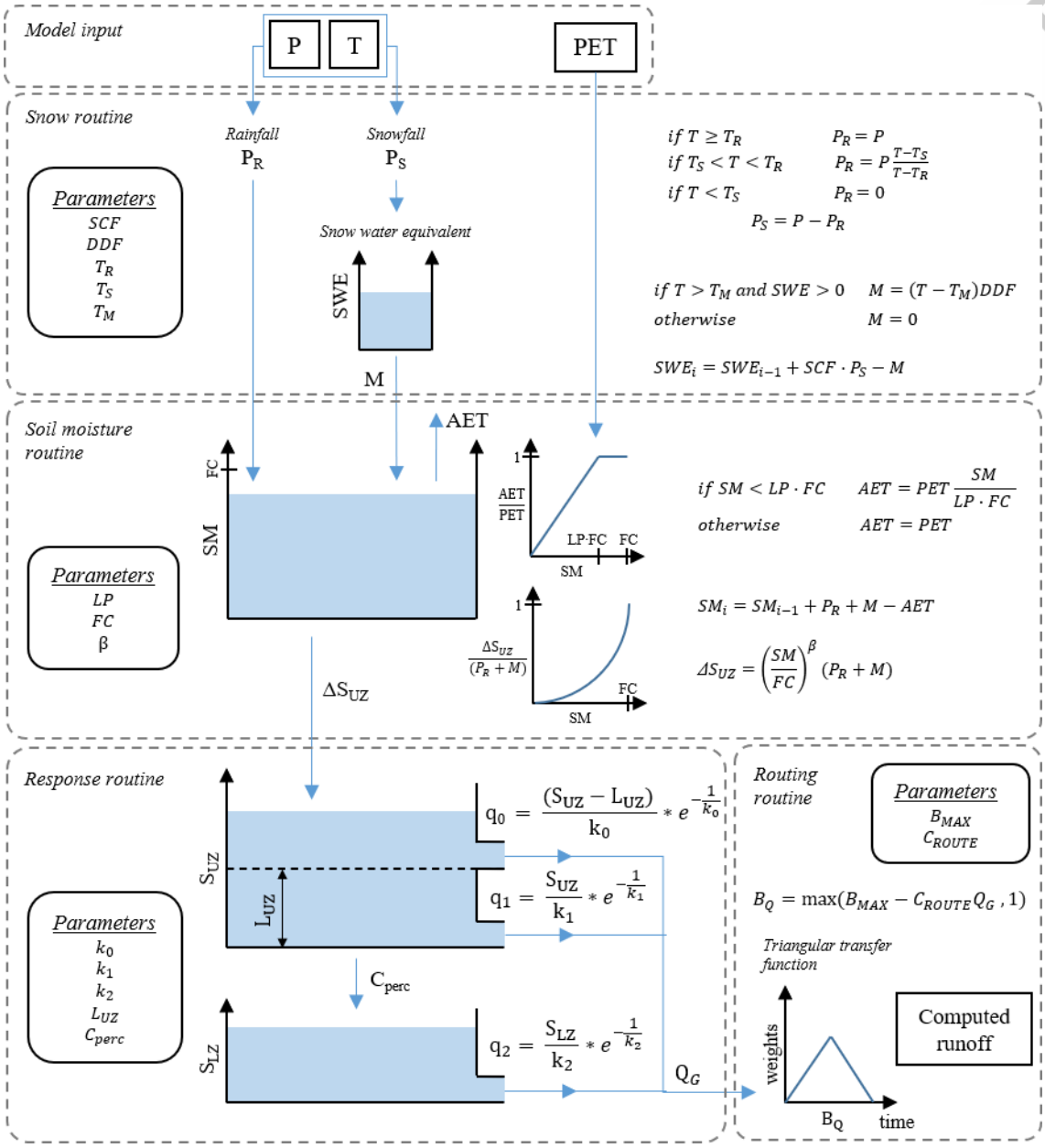
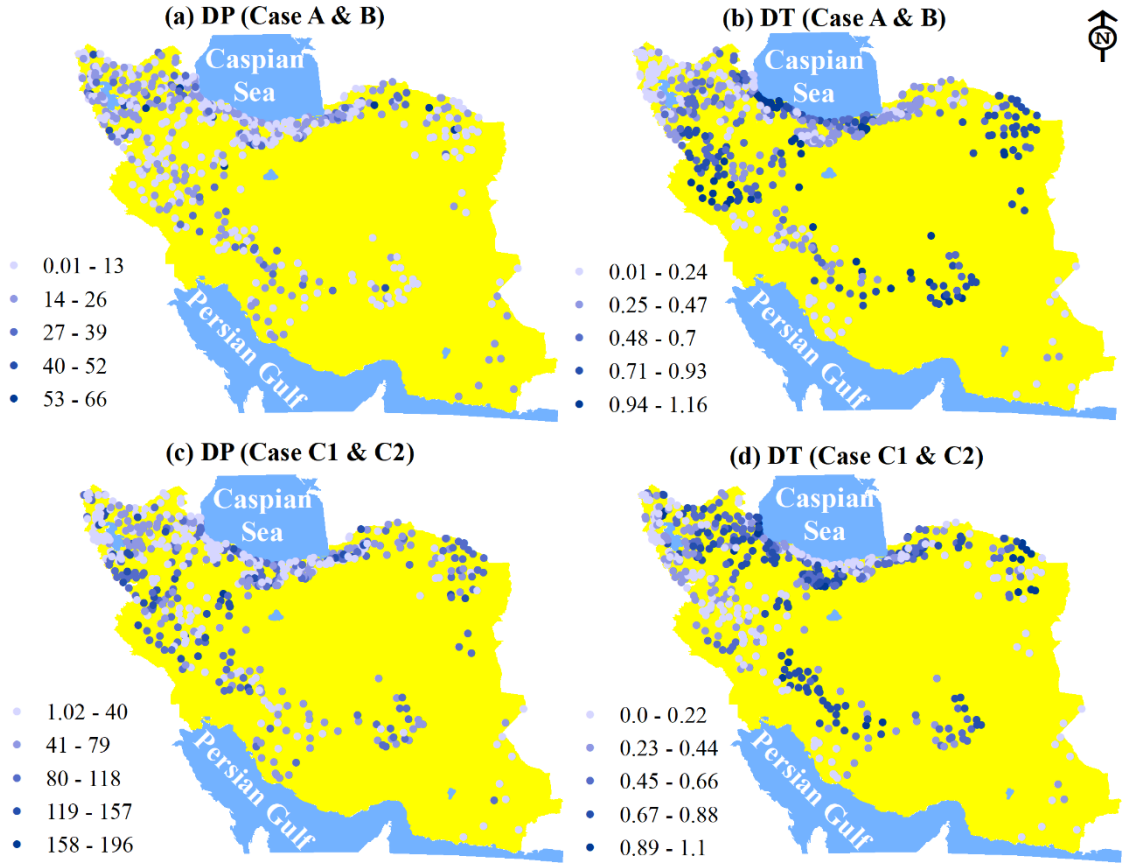


Figure C1. The structure of TUW model scheme - lumped version (from Neri et al., 2020).

Appendix D

The difference in climate variables

The spatial distribution of precipitation (DP) and temperature (DT) variability between calibration and validation periods under all eight cases.



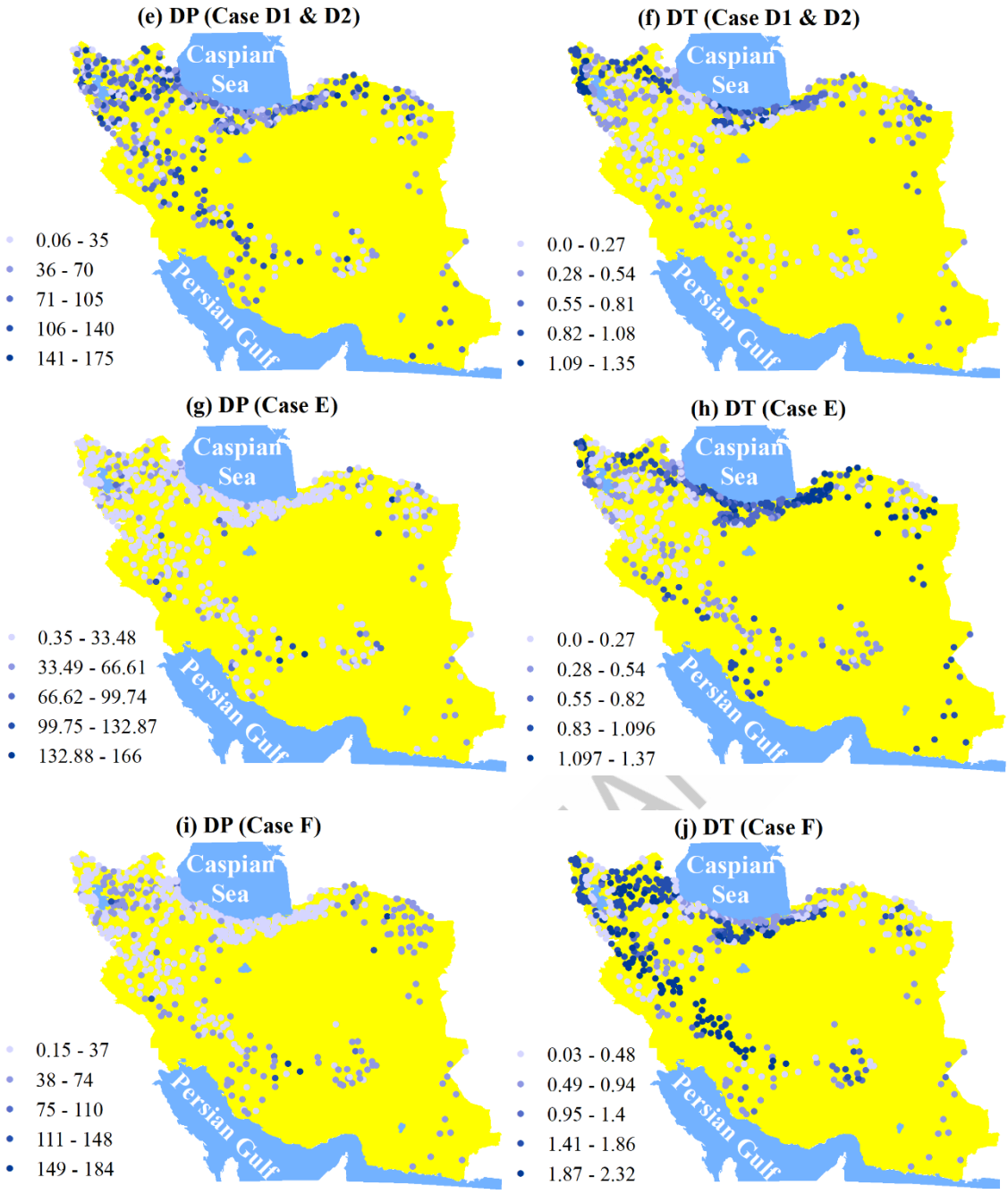


Figure D1. Spatial distribution of the difference in precipitation (DP) and temperature (DT) between calibration and validation periods.

Tables

Table 1. Catchment descriptors mean values for different climate and homogenous regions (n = 576).

Catchment descriptor	Climate region			Homogenous region		
	Snow	Warm temperate	Arid	c1	c2	c3
No. of catchments	73	235	268	146	256	174
Aridity index (-)	0.32	0.55	1.01	1.16	0.61	0.57
Area (km^2)	1025	1594	2681	3332	1953	1043
Mean elevation (m)	1500	1048	900	515	1037	1473
Mean slope (%)	24.1	23.2	17.4	16	23.4	23.9
Rangeland (%)	8.3	6.1	11.4	13.2	6.6	8.5
Agriculture (%)	12.5	14.2	15.5	14.6	14.6	14.7
Forest (%)	4.5	10.2	7.06	8.6	8.8	6.2
Residential (%)	3.1	4.1	3.7	3	4.2	3.6

Table 2. Annual mean values of the hydro-climatic attributes for the different climate and homogeneous regions from 2000 to 2014).

Climate variable	Climate region			Homogeneous region			All catchments
	Arid	Warm temperate	Snow	C1	C2	C3	
P (mm/month)	416	548	615	627	519	510	527.5
T (^{o}C)	17.2	12.3	11.5	11.3	12.8	16.4	13.5
PET (mm)	512	391	344	326	382	518	416.2
Runoff (mm)	51	74	86	81	66	61	69.3

Table 3. The HBV model's calibrated parameters, their value ranges (lower and upper limits), correlation coefficient (r_1), and median values for two calibration periods.

Parameter	Description	Lower	Upper	Median	r_1
SCF	Snow correction factor [-]	0.9	1.5	1.17	0.57
DDF	Degree day factor [mm/^{o}C day]	0	5	2.41	0.61
T_R	Threshold temperature above which precipitation is rain [^{o}C]	1	3	1.82	0.48
T_S	Threshold temperature below which precipitation is snow [^{o}C]	-3	1.0	-0.63	0.57
T_M	Threshold temperature above which melt starts [^{o}C]	-2	2	0.05	0.44
LP	Parameter related to the limit for potential evaporation [-]	0	1	0.41	0.43
FC	Field capacity [mm]	0	600	301.56	0.57
β	The nonlinear parameter for runoff production [-]	0	20	10.18	0.41
K_0	Storage coefficient for very fast response [day]	0	2	0.82	0.47
K_1	Storage coefficient for fast response [day]	2	30	15.1	0.53
K_2	Storage coefficient for slow response [day]	30	250	137.26	0.31
l_{UZ}	Threshold storage state	0	100	46.21	0.49
Cperc	Constant percolation rate [mm/day]	0	8	3.29	0.38
B_{MAX}	Maximum base at low flows [day]	0	30	14.14	0.45
C_{ROUTE}	Free scaling parameter [day^2 /mm]	0	50	19.88	0.32

Figure captions:

Figure 1. Map of Iran with the selected streamflow gauges belong to climate regions (a) and homogenous regions (b).

Figure 2. Variability in mean monthly (left panel) and annual (right panel) precipitation, temperature, PET, and runoff over Iranian catchments for three climate regions and the entire country (Iran). The long-term mean is indicated by dashed line.

Figure 3. Climatic classification of the water years for all study catchments from 2000 to 2014 period. Annual mean precipitation (mm) and temperature (^{o}C) values for the calibration and validation are included in parenthesis. Dry and wet conditions are indicated by yellow and blue colors, respectively.

Figure 4. A detailed step-by-step implementation of all processes and model transferability.

Figure 5. Boxplot of calibration and model transferability under classical and differential split-sample tests. The dashed lines at the top (NSE = 0.75), middle (NSE = 0.65), and bottom (NSE = 0.5) indicate the criteria of Moriasi et al. (2007).

Figure 6. A 1:1 plot of all 15 HBV model parameter values for each of the calibration periods at climate regions. The legend for climate regions indicated in Fig. 1a.

Figure 7. A 1:1 plot of all 15 HBV model parameter values for each of the calibration periods at the homogeneous regions. The legend for homogenous regions indicated in Fig. 1b.

Figure 8. A 1:1 plot of the difference in model parameters (ΔParameter) and (a to h) difference in precipitation (ΔP), (i to p) difference in temperature (ΔT). The legend for climate regions indicated in Fig. 1a.

Figure 9. A 1:1 plot of the difference in model parameters (ΔParameter) and (a to h) difference in precipitation (ΔP), (i to p) difference in temperature (ΔT). The legend for homogenous regions indicated in Fig. 1b.

Figure 10. Boxplot of calibration and model transferability for climate (a) and homogeneous regions (b). The dashed lines at top (NSE = 0.75), middle (NSE = 0.65), and bottom (NSE = 0.55) indicate the criteria of Moriasi et al. (2007). c1, c2, and c3 correspond to homogenous regions clusters 1, 2 and 3, respectively.

Figure 11. (a) Benchmark NSE(Q) score that the HBV model must beat to be considered acceptable in each catchment. (b) Benchmark type (median or mean calendar day flow).

Figure 12. NSE mean values for case A and case B (a and c). The location of the catchments where the model outperformed the benchmark score (b and d). Difference between the validation NSE and the benchmark score in each catchment (e and f).

Figure 13. Mean values of NSE over differential split sample-test (left panel). The location of the catchments where the model outperformed the benchmark score (right panel).

Figure 14. Map of catchment locations where the temporal or spatial parameter transfer method is the best performing mode over classical and differential split-sample tests.

Figure 15. Correlation matrix between catchment descriptors and difference in NSE values of calibration and validation periods (DVAL) (local scale). Differences in precipitation and temperature values between the calibration and validation periods are denoted by DP and DT, respectively.

Figure 16. Correlation matrix between catchment descriptors and difference in NSE values of calibration and regionalization methods (local scale). For physical similarity and spatial proximity, DPS and DSP are differences in NSE values between calibration and spatial modes.

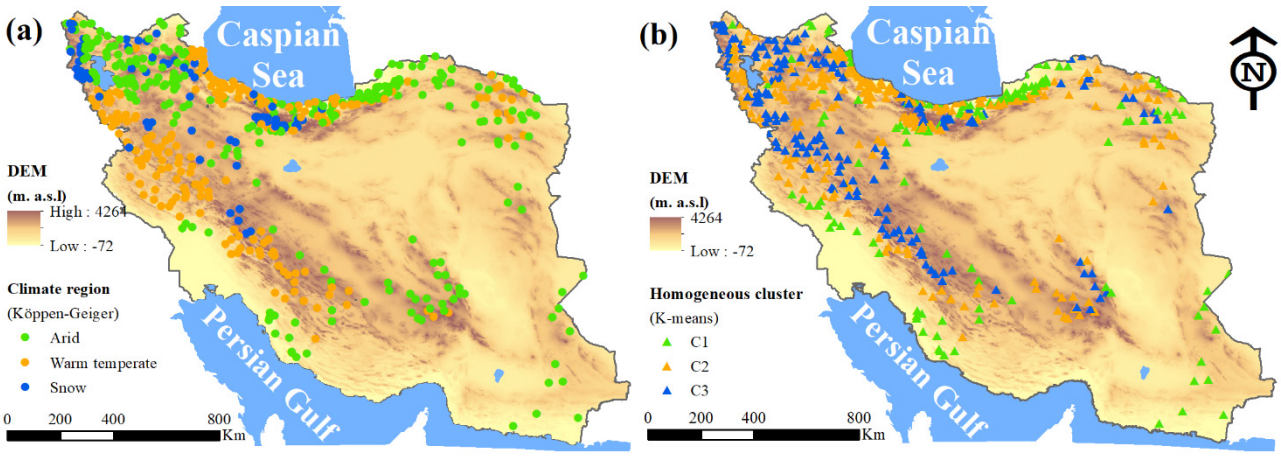


Figure 1.

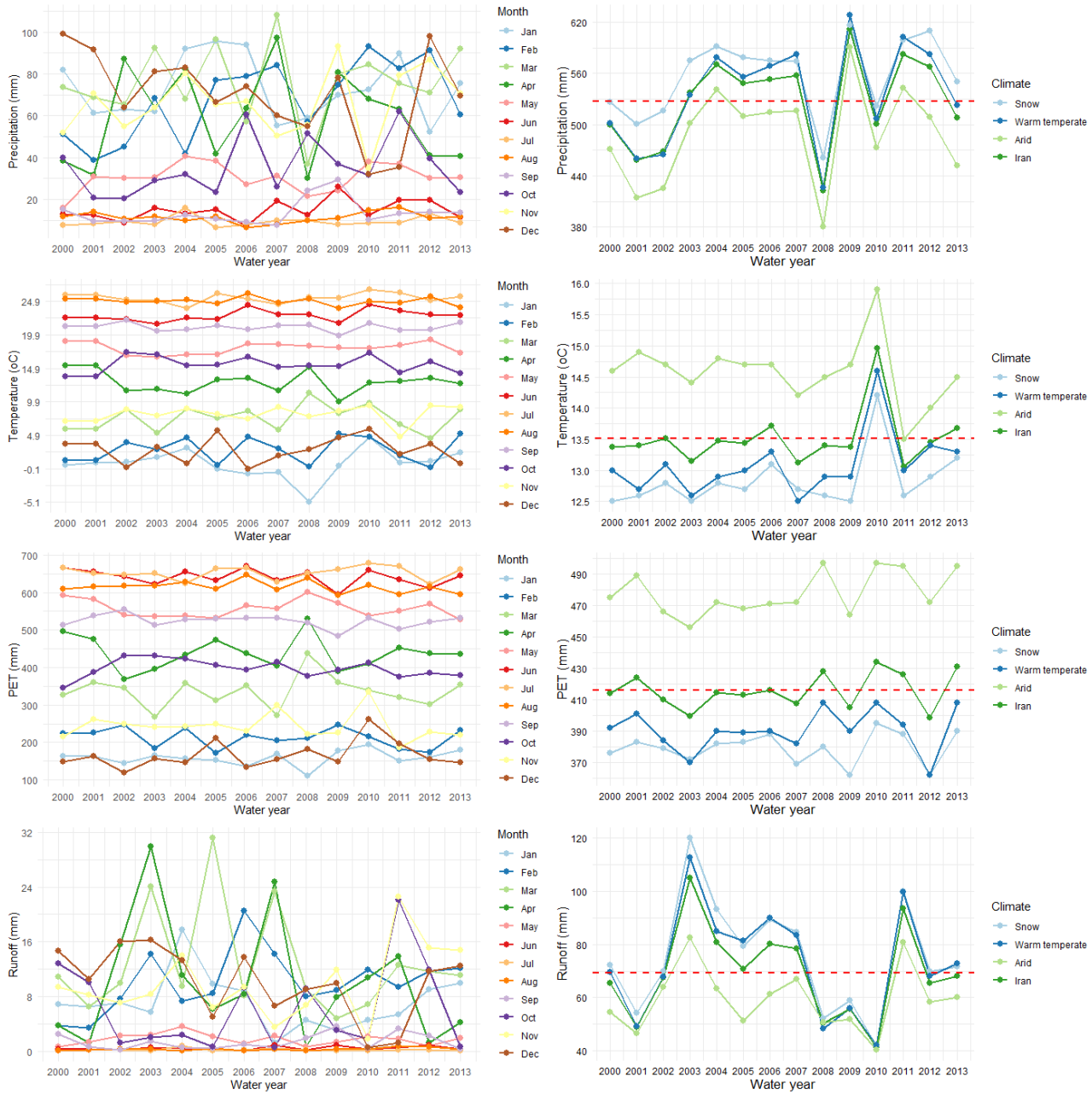


Figure 2.

Case	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
A (CSST)	Calibration (519/13.4)							Validation (536/13.5)						
B (CSST)	Validation							Calibration						
C1 (DSST)	Calibration (491/13.3)				Validation (557/13.5)			-						
C2 (DSST)	Validation				Calibration			-						
D1 (DSST)	-							Calibration (523/13.7)				Validation (553/13.3)		
D2 (DSST)	-							Validation				Calibration		
E (DSST)	Calibration (507/13.3)						-			Validation (534/13.7)				
F (DSST)	Calibration (475/13.4)						-					Validation (552/13.3)		

Figure 3.

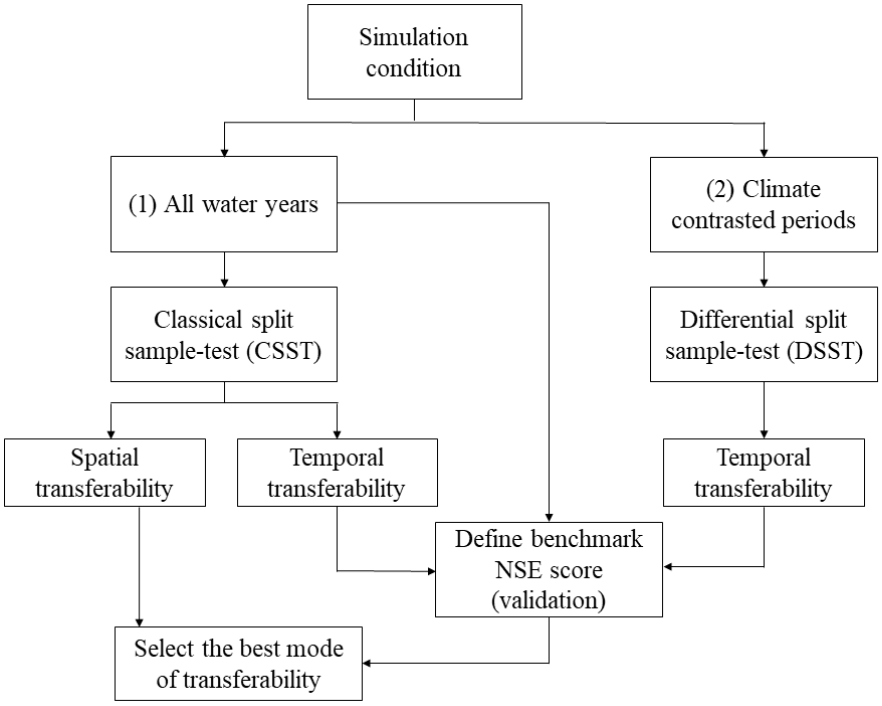


Figure 4.

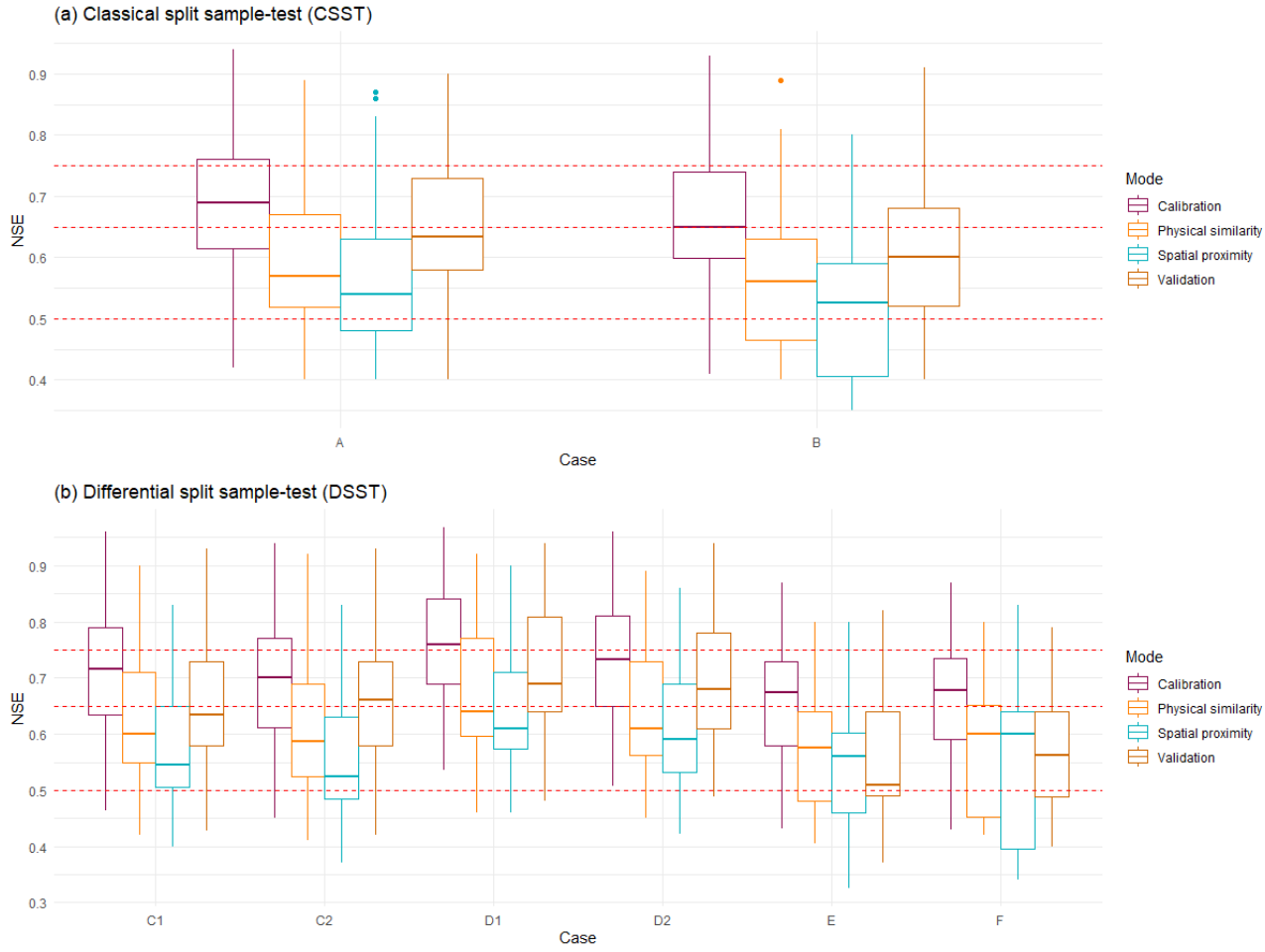


Figure 5.

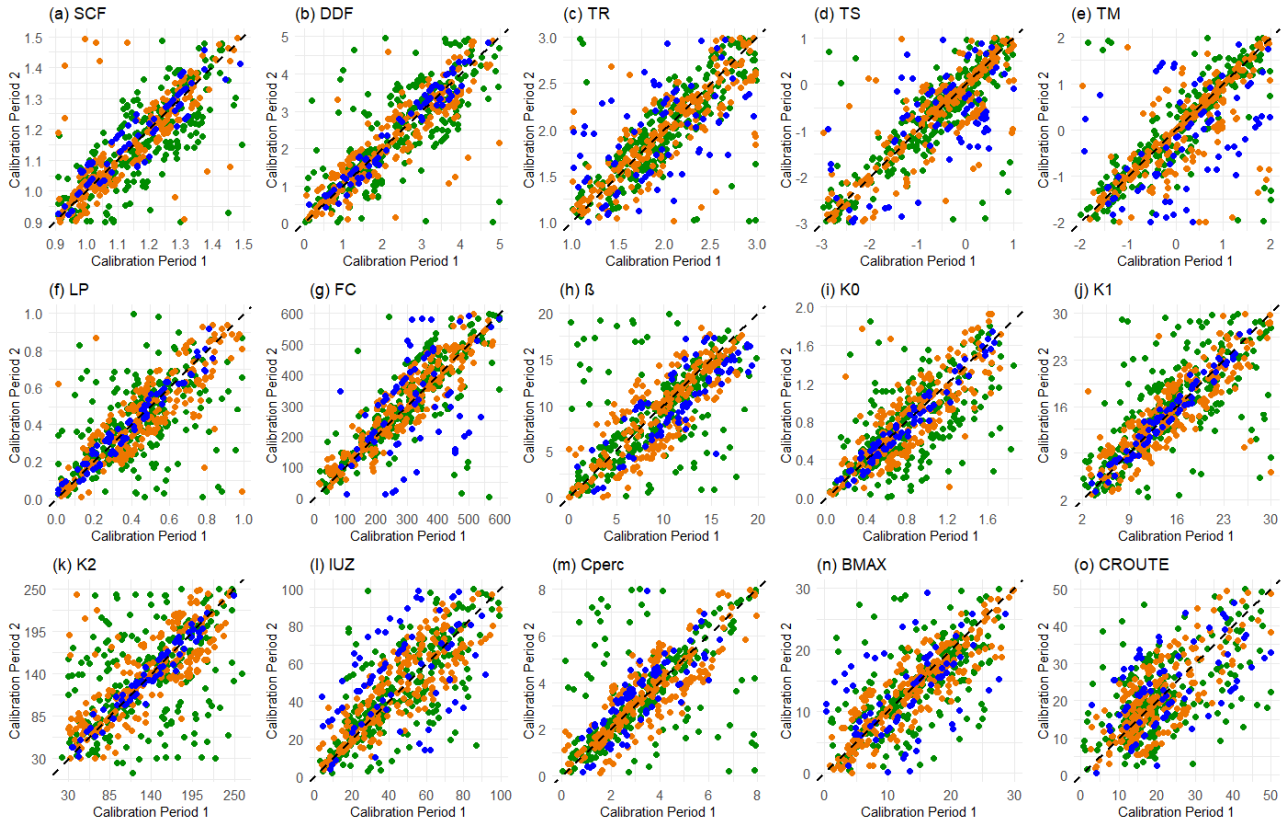


Figure 6.

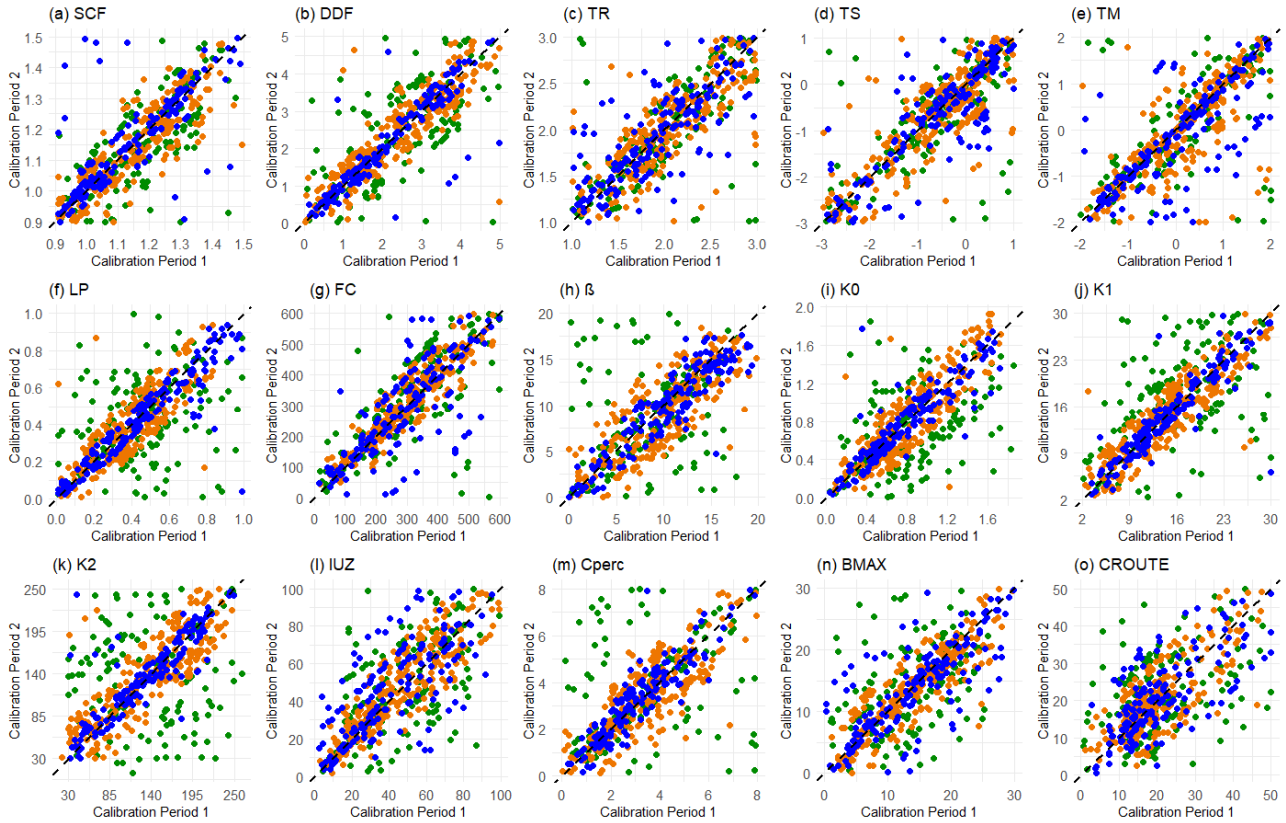


Figure 7.

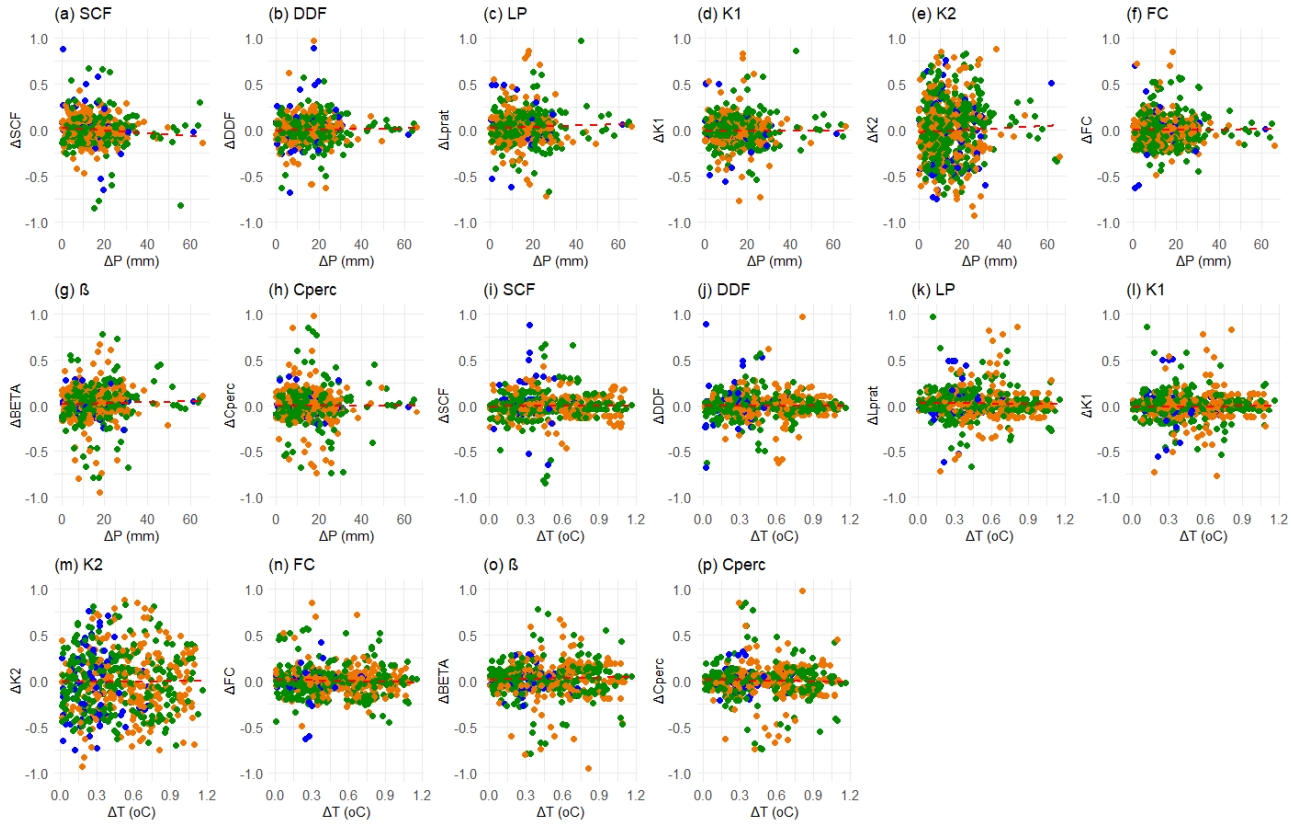


Figure 8.

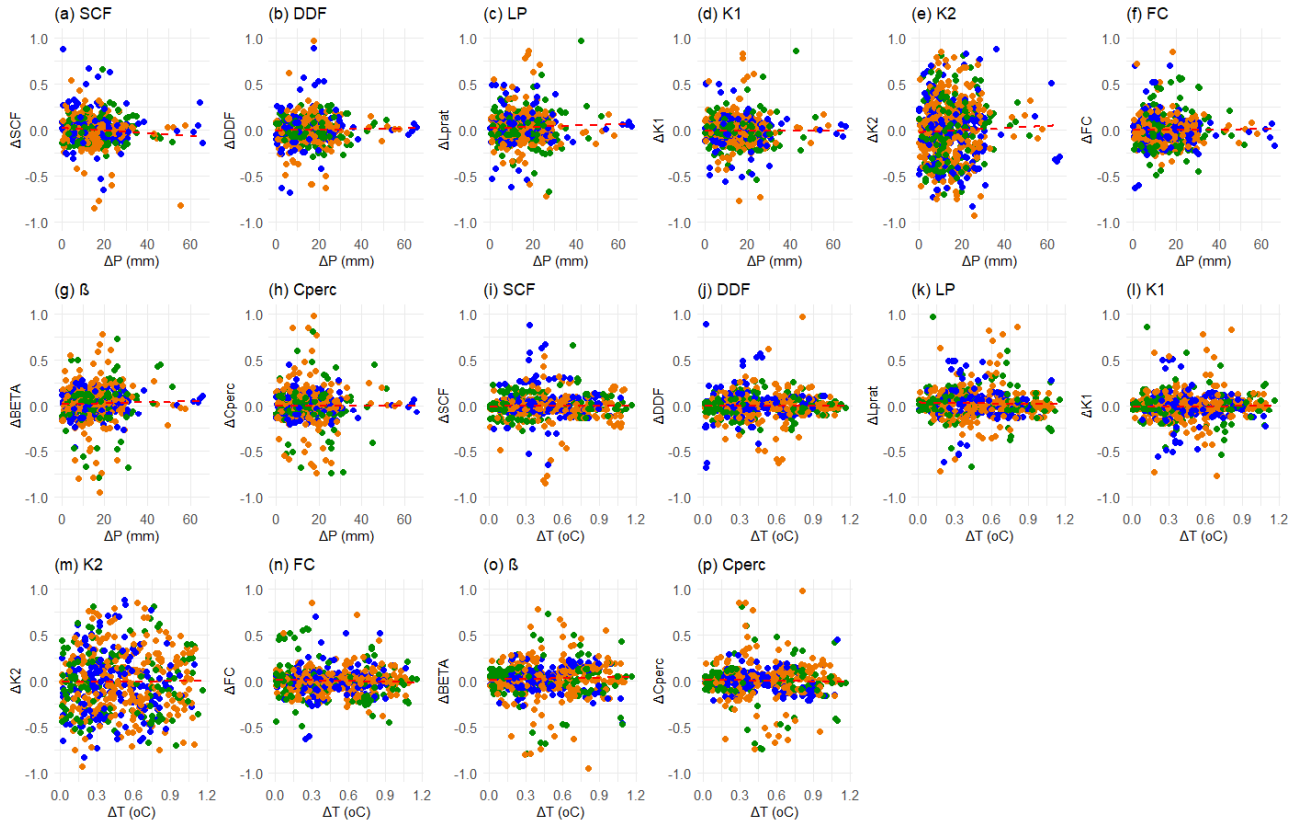


Figure 9.

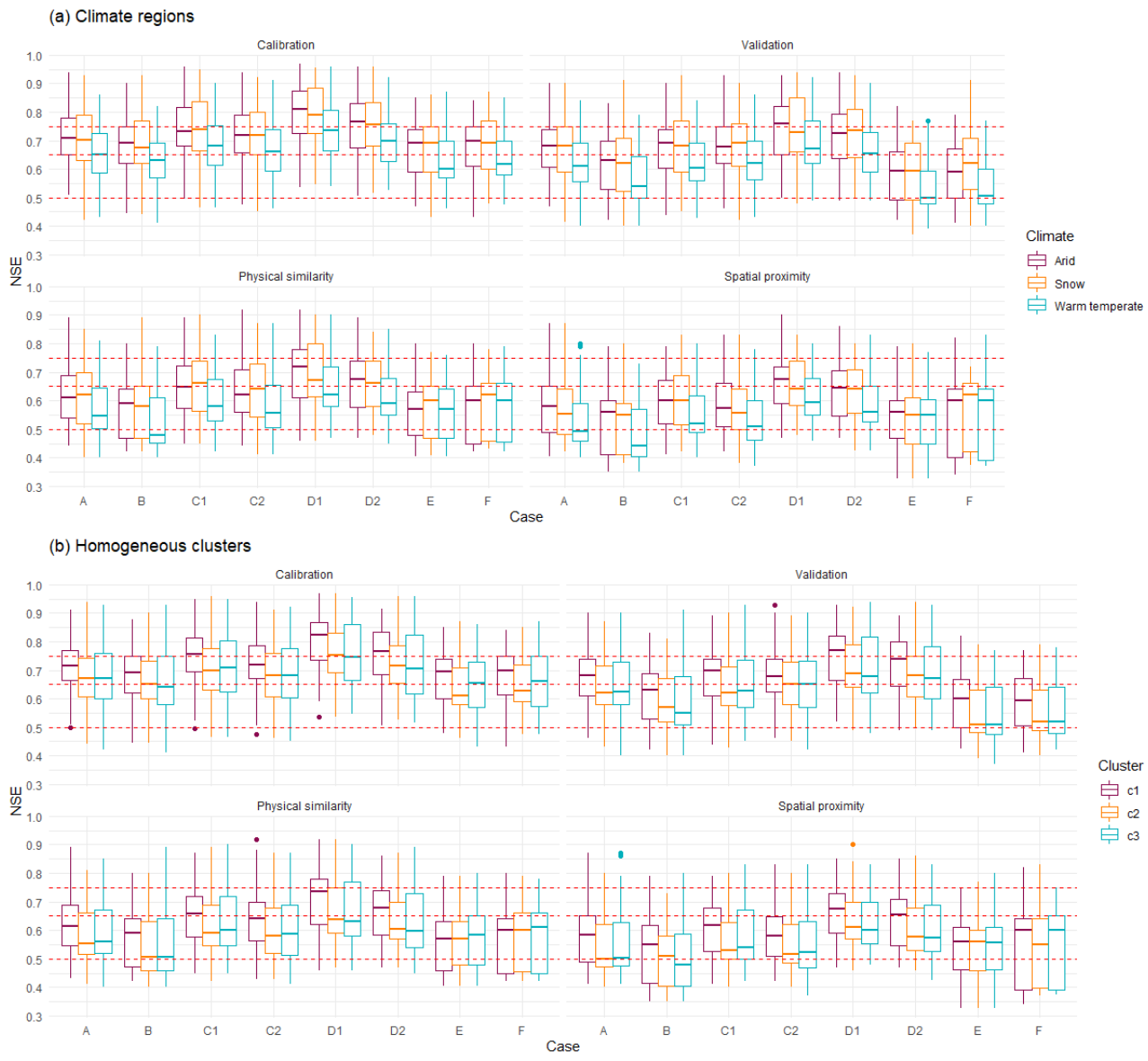


Figure 10.

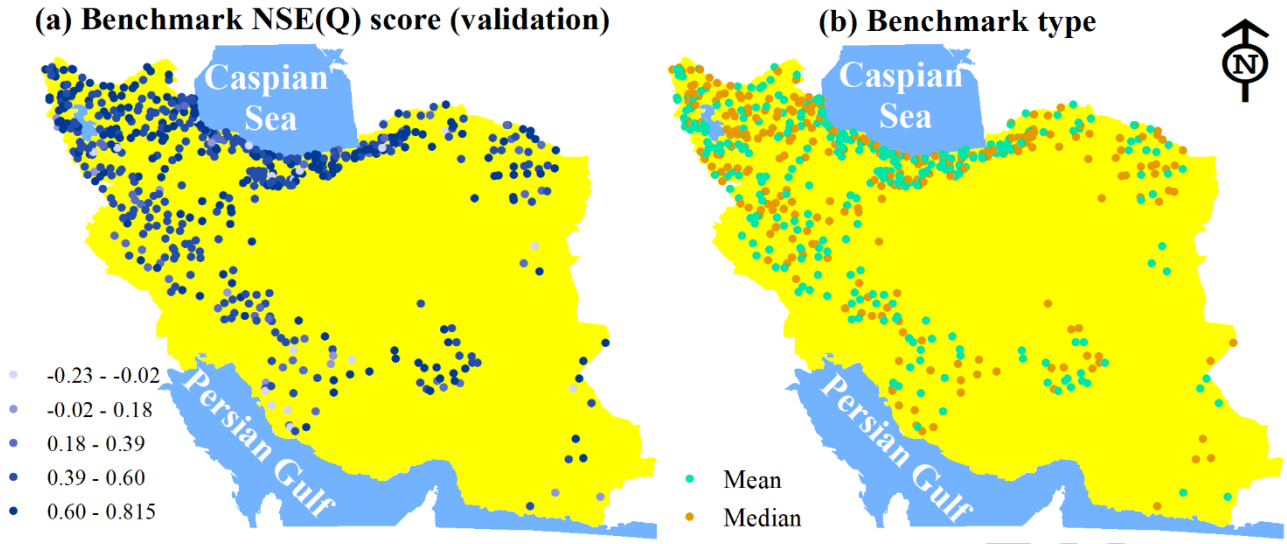


Figure 11.

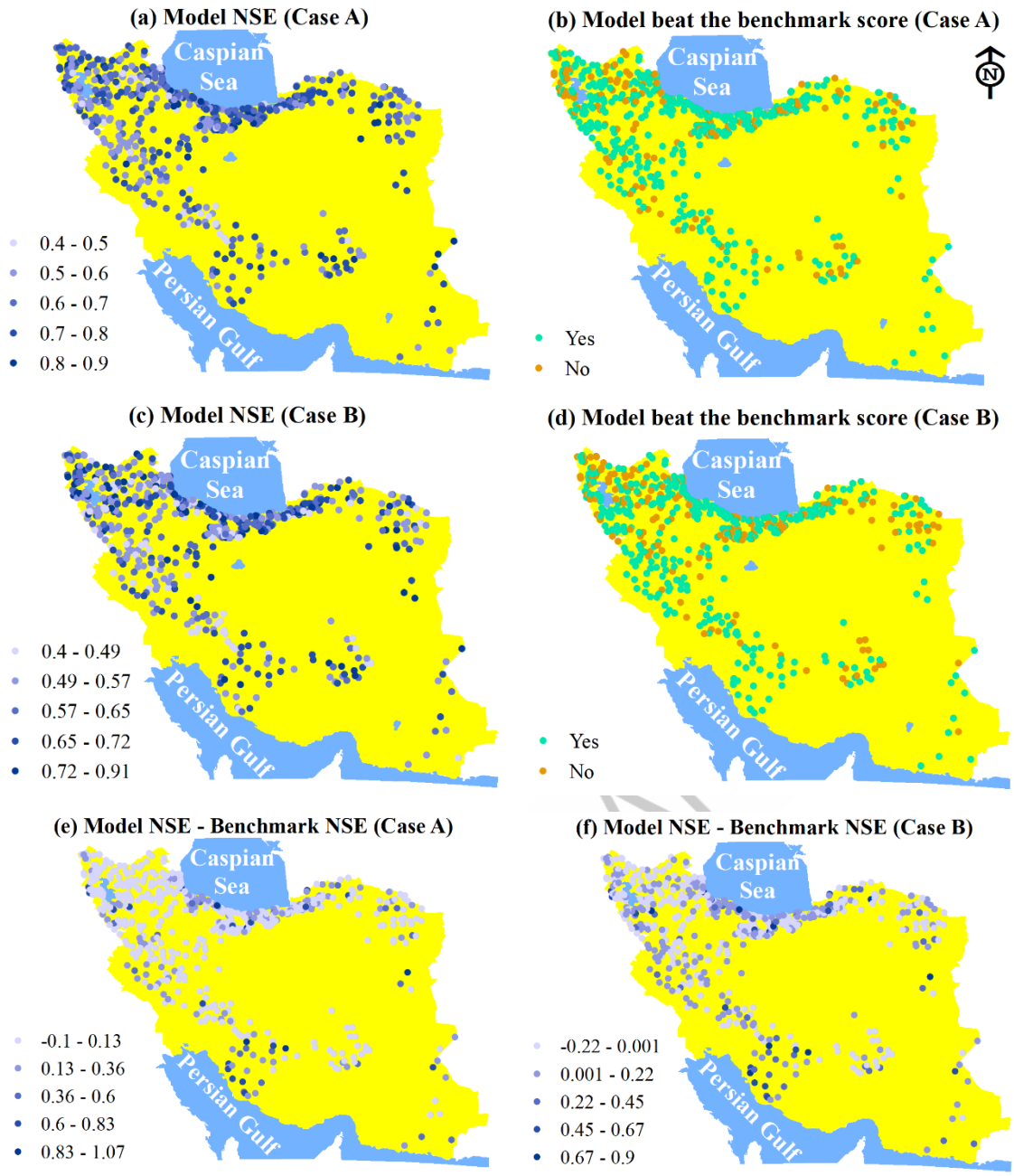
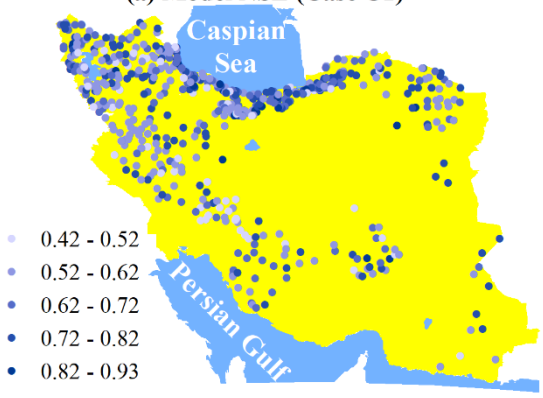
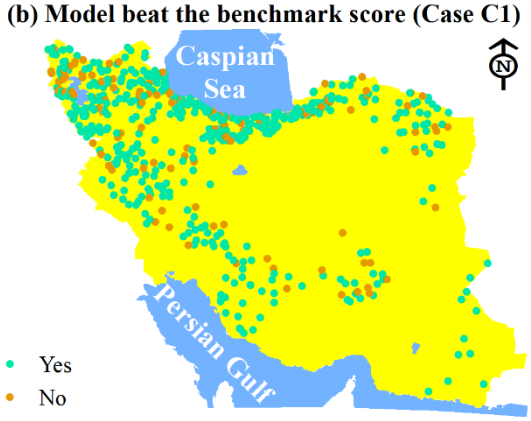


Figure 12.

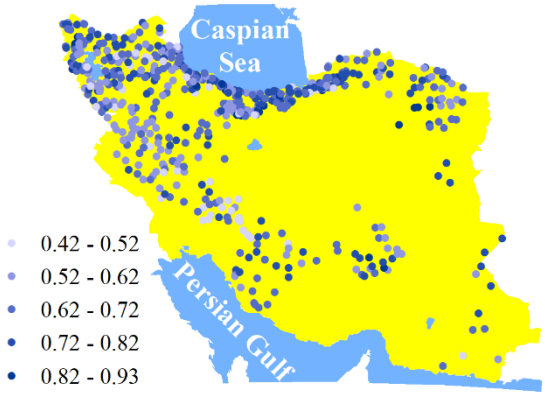
(a) Model NSE (Case C1)



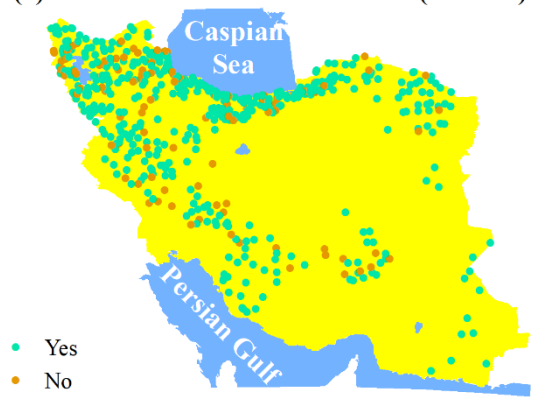
(b) Model beat the benchmark score (Case C1)



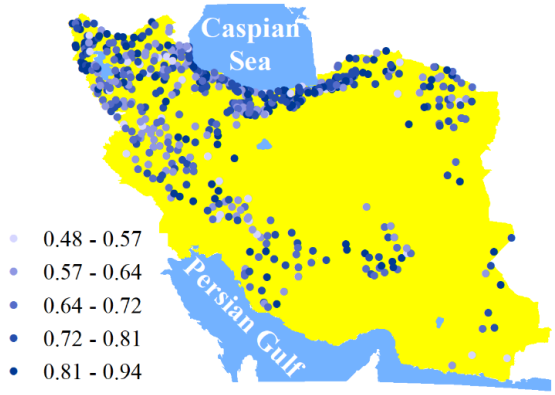
(c) Model NSE (Case C2)



(d) Model beat the benchmark score (Case C2)

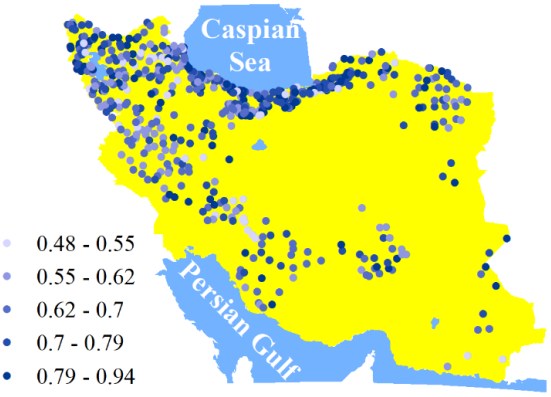


(e) Model NSE (Case D1)

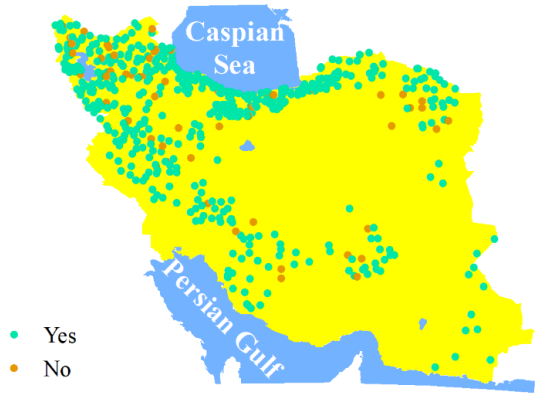


(f) Model beat the benchmark score (Case D1)

(g) Model NSE (Case D2)



(h) Model beat the benchmark score (Case D2)



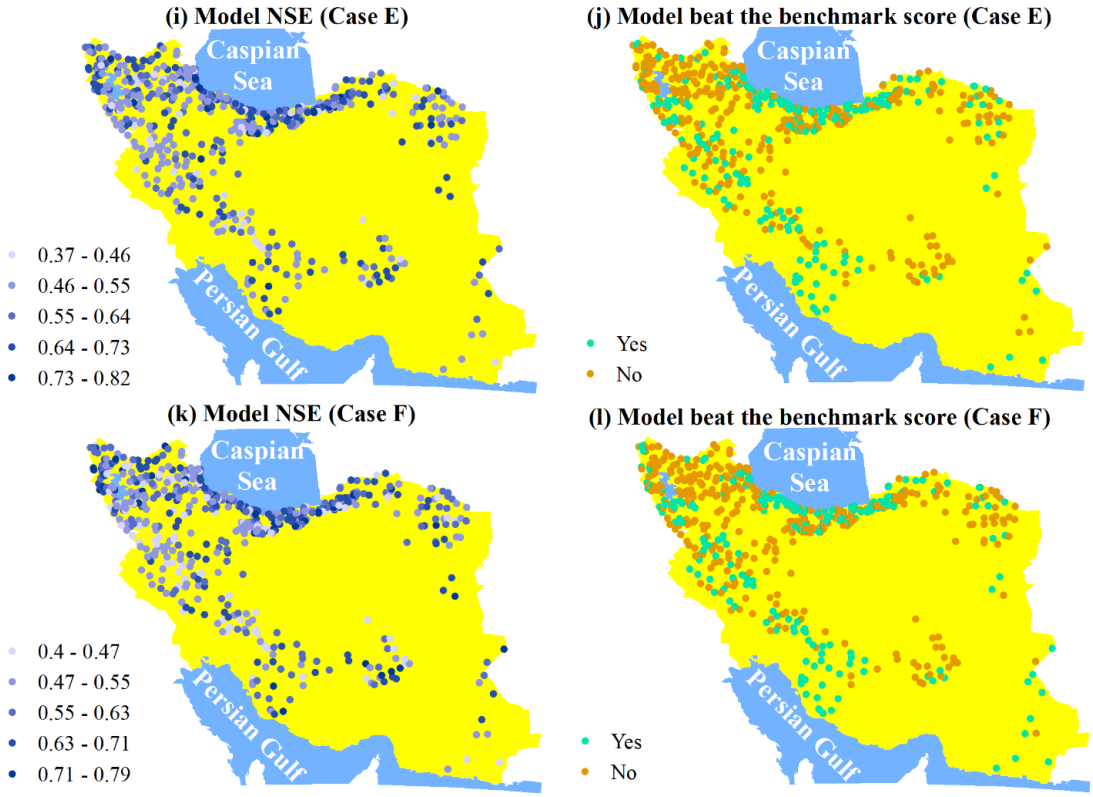


Figure 13.

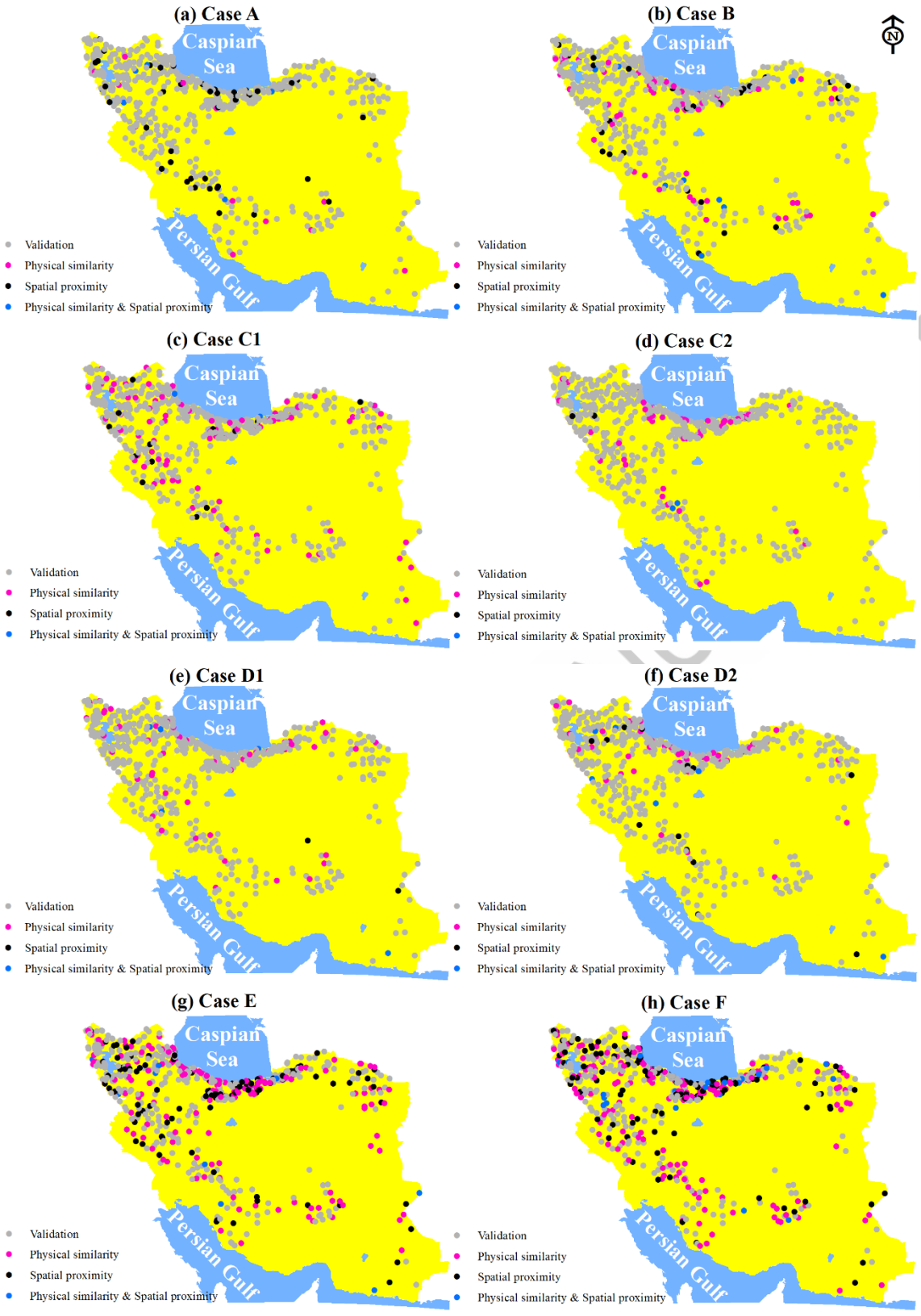

(a) Case A
(b) Case B
(c) Case C1
(d) Case C2
(e) Case D1
(f) Case D2
(g) Case E
(h) Case F
Caspian Sea
Persian Gulf
Validation
Physical similarity
Spatial proximity
Physical similarity & Spatial proximity
N

Figure 14.

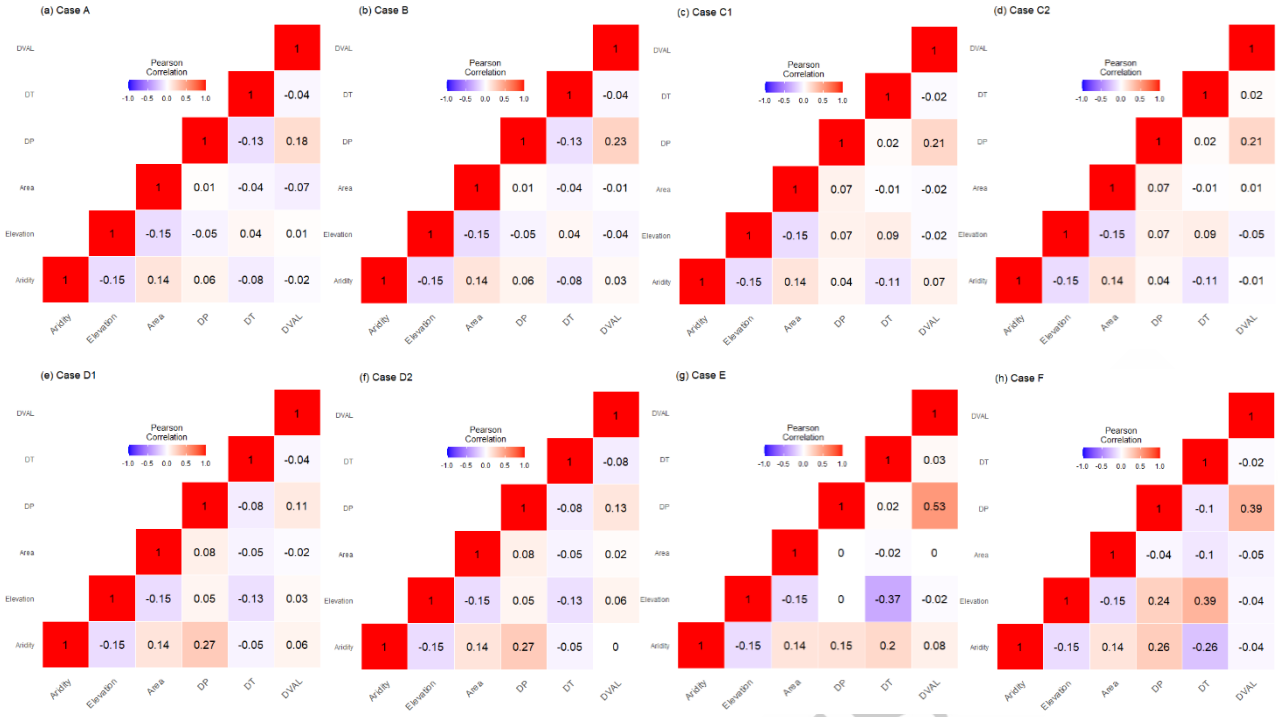


Figure 15.

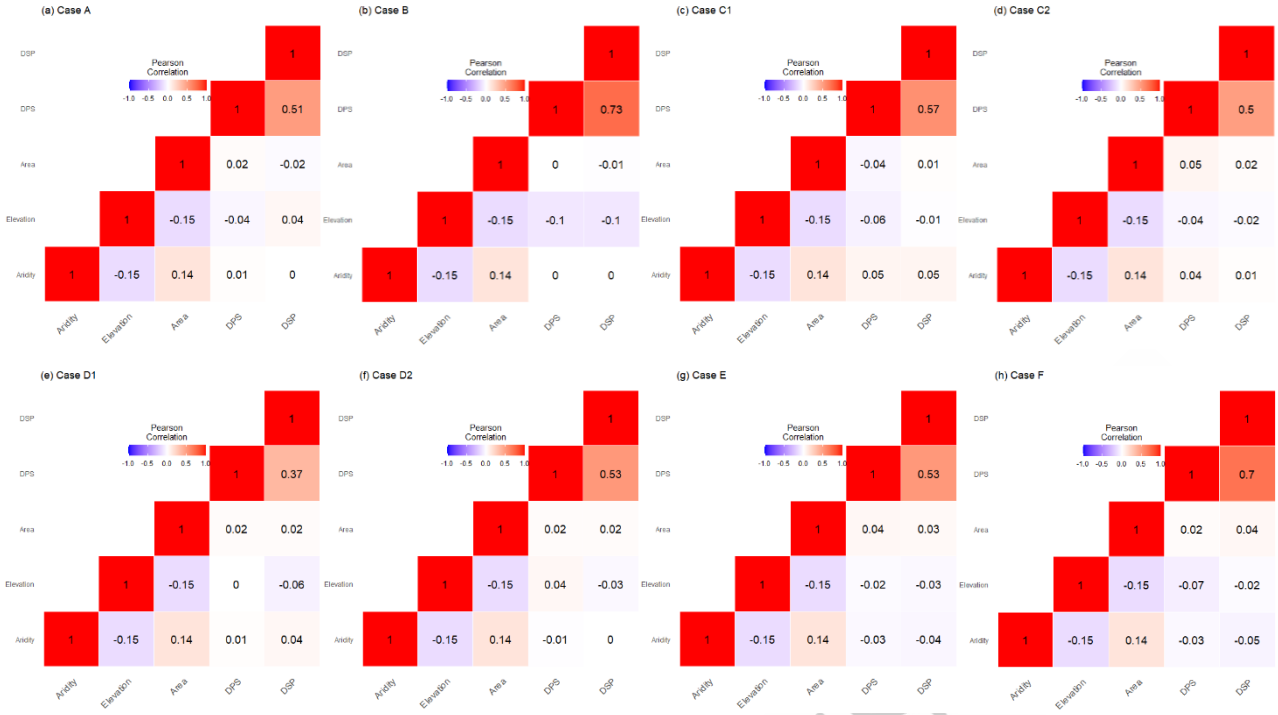


Figure 16.