Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Research papers

Dependence of regionalization methods on the complexity of hydrological models in multiple climatic regions

Xue Yang[a], Jan Magnusson[b], Shaochun Huang[b], Stein Beldring[b], Chong-Yu Xu[a,*]

[a] Department of Geosciences, University of Oslo, P.O. Box 1047 Blindern, 0316 Oslo, Norway
[b] Norwegian Water Resources and Energy Directorate (NVE), P.O. Box 5091 Majorstua, 0301 Oslo, Norway

ARTICLE INFO

Keywords:
Regionalization methods
Hydrological models
Climatic zones
Ungauged catchments

ABSTRACT

Hydrological models have been widely used to predict runoff in regions with observed discharge data, and regionalization methods have been extensively discussed for providing runoff predictions in ungauged basins (PUB), especially during the PUB decade (2003–2012). Great progress has been achieved in the field of regionalization in previous studies, in which different hydrological models have been coupled with various regionalization methods. However, different conclusions have been drawn due to the use of different hydrological models, regionalization methods, and study regions. In this study, we assessed the performance of the five most widely used regionalization methods (spatial proximity with parameter averaging option (SP-par), spatial proximity with output averaging option (SP-out), physical similarity with parameter averaging option (Phy-par), physical similarity with output averaging option (Phy-out), and regression methods (PCR)) and four daily rainfall-runoff models (GR4J, WASMOD, HBV and XAJ, with 6, 8, 13, and 17 parameters, respectively) at the same time. Our aim was to evaluate how the performance of the regionalization methods depends on (a) the selection of hydrological models, (b) nonstationary climate conditions, and (c) different climatic regions. This investigation used data from 86 independent catchments evenly distributed throughout Norway, covering three different climate zones (oceanic, continental and polar tundra) according to the Köppen-Geiger classification. The results showed that (a) the SP-out and Phy-out methods performed better than the SP-par and Phy-par for all the hydrological models, and the regression method performed worst in most cases; (b) the difference between the parameter averaging option and the output averaging option is positively related to the number of hydrological model parameters, i.e. the greater the number of parameters, the larger the difference between the two options; (c) the XAJ model with the greatest number of parameters produced the best results in most cases, and models with fewer parameters tend to produce similar performance for the different regionalization methods; (d) models with more parameters displayed larger declines in performance than those with fewer parameters for nonstationary conditions; and (e) clear differences in the performance of the regionalization methods exist among the three climatic regions. This study provides insight into the relationship between the complexity of hydrological models and regionalization methods in cold and seasonally snow-covered regions.

1. Introduction

Runoff prediction plays a significant and essential role in water resources management, the assessment of the impact of environmental change (e.g., climate and land use), and hydrological design (e.g., Blöschl and Montanari, 2010; Parajka et al., 2013). During the last several decades, hydrological models have become the most popular and common solution for runoff predictions. However, the models have free parameters to be calibrated by using the observed discharge data before predicting the runoff hydrographs, which are not available in many catchments of interest (e.g., He et al., 2011; Parajka et al., 2013). This fact made the topic 'predictions in basins without observed discharge data (ungauged basins)' attractive and challenging for hydrologists (e.g., Parajka et al., 2007; Sivapalan et al., 2003; Xu, 2003). As a result, the International Association of Hydrological Sciences (IAHS) established a "Decade on Predictions in Ungauged Basins (PUB): 2003–2012", and great progress has been achieved during this period (Hrachowitz et al., 2013).

Regionalization is defined as the method for predicting runoff in ungauged basins by transferring information from gauged (donor) to ungauged (target) catchments (e.g., Rojas-Serna et al., 2016; Razavi and Coulibaly, 2013). In general, regionalization methods are classified

[*] Corresponding author.
E-mail address: c.y.xu@geo.uio.no (C.-Y. Xu).

https://doi.org/10.1016/j.jhydrol.2019.124357
Received 17 April 2019; Received in revised form 6 November 2019; Accepted 12 November 2019
Available online 15 November 2019

into three categories: (a) spatial proximity methods assume that geographically close catchments have similar hydrological behaviors (e.g., Egbuniwe and Todd, 1976; Vandewiele et al., 1991); (b) physical similarity methods assume that catchments with similar physical characteristics have the same hydrological response (e.g., Burn and Boorman, 1993; McIntyre et al., 2005), thus, the parameter values are transferred to ungauged basins from either geographically close or physically similar gauged basins; and (c) the regression method, which is one of the most popular and oldest regionalization approaches (Oudin et al., 2008), links model parameters to physical and climatic catchment characteristics by regression functions and assumes that the relationship is transferable from gauged to ungauged basins (e.g., Magette et al., 1976; Young, 2006).

Many studies have applied and compared regionalization methods for various regions in combination with a wide range of hydrological models. However, in many cases, the conclusion about which method performed best differs largely among the studies. For example, Merz and Blöschl (2004) concluded that the spatial proximity method performed better than the regression method for catchments in Austria using the HBV model. On the other hand, Young (2006) found that the regression method gave better results than the spatial proximity method in the UK. Bao et al. (2012) concluded that the physical similarity method was best by using the Akaike information criterion (AIC) on 55 catchments in China. Different models were applied for different regions in these studies, and therefore many hydrologists claim that the performance of regionalization methods depends on the study area and the choice of hydrological model (e.g., Parajka et al., 2013; Reichl et al., 2009; Salinas et al., 2013; Samuel et al., 2011; Viglione et al., 2013). Most of the above-mentioned studies only used one hydrological model in a specific region, and conclusions cannot be drawn on how the model selection or study region affects the performance of the regionalization methods.

Few studies have assessed the performance of regionalization methods using multiple models. Li and Zhang (2017) used SIMHYD (10 model parameters) and XAJ (12 model parameters) in Australia and found consistent regionalization results for both models. The same conclusion was drawn by Li et al. (2014), where GR4J (7 model parameters) and SIMHYD (12 model parameters) were applied in the southeast Tibetan Plateau. Furthermore, Petheram et al. (2012) conducted a comparison by using five rainfall-runoff models and concluded that the difference between hydrological models was negligible for runoff prediction in ungauged basins. This conclusion was consistent with two other studies (Chiew, 2010; Viney et al., 2009b), which also included five hydrological models. However, none of these studies included a regression approach, which provided very different results when used with either the GR4J (4 model parameters) or TOPMO (6 model parameters) model in the study of Oudin et al. (2008), who tested three kinds of regionalization methods using two hydrological models for 913 catchments in France. Either the number of regionalization methods or the number of models used in previous studies is still too small to draw a general conclusion. In addition, all these evaluations have been performed for relatively warm climate regions, where the snow process is of limited importance. Thus, a more comprehensive study is needed to investigate how regionalization performance differs with multiple hydrological models of different complexity for runoff prediction in ungauged basins, especially for cold and seasonally snow-covered regions.

Furthermore, climate is changing (IPCC, 2014), resulting in nonstationary relationships between rainfall and runoff (Zhang et al., 2011), which makes the reliability of applying the conclusions made in a historical period into future application questionable. Thus, for future runoff prediction in ungauged basins, it is essential to investigate the transferability of the regionalization methods under changing climatic conditions (e.g., Broderick, 2016; Yang et al., 2019). Finally, regionalization performances also vary between regions, according to Parajka et al. (2013), who statistically summarized this conclusion from 34 regionalization studies. However, it cannot explicitly present the performance difference between regions for specifically selected regionalization methods because different hydrological models and regionalization methods were applied in the studies cited and summarized by Parajka et al. (2013).

In this study, we perform a comprehensive evaluation of the performance of five widely used regionalization methods (see Section 3.2) combined with four frequently used hydrological models (GR4J–6 parameters, WASMOD–8 parameters, HBV–13 parameters and XAJ–17 parameters) in regions with highly contrasting physiographic and climatic settings. The evaluation is based on 86 catchments in Norway, belonging to three different climatic regions according to the Köppen-Geiger classification (Kottek et al. 2006) and under different climate conditions. This is the first study that specifically addresses how the performance of the regionalization methods (a) depends on the selection of hydrological models, (b) changes in different climate conditions, i.e., when air temperature increases, and (c) varies between different climate regions as defined by the Köppen-Geiger classification.

2. Study area and data

2.1. Study area

Our study catchments are located in Norway, which is situated in northern Europe in the western and northern part of the Scandinavian Peninsula. Norway has a long and rugged coastline, elevation spanning from sea level to 2469 m.a.sl., and latitudes ranging from 58° to 71°N. This results in highly variable hydroclimatological conditions across the study domain (Vormoor et al., 2016; Yang et al., 2018, 2019). In this study, we used data from 86 nonoverlapping catchments distributed evenly throughout our study domain (Figure 1). These stations have continuous meteorological data and discharge data records with less than 40% missing values during the periods from 1980 to 1989 as well as 2006 to 2015. These two periods are used in this study. The left panel map in Figure 1 also displays the Köppen-Geiger climate classification, which is based on data from 1976 to 2000 (Kottek et al., 2006; Peel et al., 2007; Beck et al., 2018). Note that the original classification divided Norway into five different climate groups. However, in two of these groups, less than 10 catchments were located. We therefore merged some of the groups, resulting in the following three regions: (a) oceanic climate containing 19 catchments, (b) continental climate containing 52 catchments and (c) polar tundra climate containing 15 catchments.

2.2. Data

For the hydrological simulations, we used daily precipitation and temperature data acquired from the gridded seNorge dataset with a resolution of 1 km produced by the Norwegian Meteorological Institute (Tveito et al., 2005; Mohr, 2009; Jansson, 2007). Daily discharge data were obtained from the hydrometric observation network of the Norwegian Water Resources and Energy Directorate (NVE). To test the performance of the regionalization methods under varying climate conditions, we analyzed the precipitation and temperature records for the period from 1980 to 2015 (Figure 2). For precipitation, there is no clear trend, whereas temperature increases throughout the study period. For model calibration and verification, we selected ten years at the start (1980 to 1989) and the end (2006 to 2015) of the whole period since these two periods show the largest difference in air temperature. For the first period, the average precipitation is 1932 mm/year, and the air temperature is 1.2 °C. For the second period, the average precipitation is 2027 mm/year, and the air temperature is 2.6 °C. The right panels in Figure 2 show the average monthly precipitation, temperature and Pardé coefficient (ratio between the average monthly discharge and the mean annual runoff) for the catchments in each climatic group. The oceanic climate group is characterized by higher precipitation during autumn and winter and higher air temperature than that of the two remaining groups. The watersheds in the oceanic climate group also show two peaks in runoff (compare the Pardé coefficient between the groups) resulting from spring snowmelt and strong

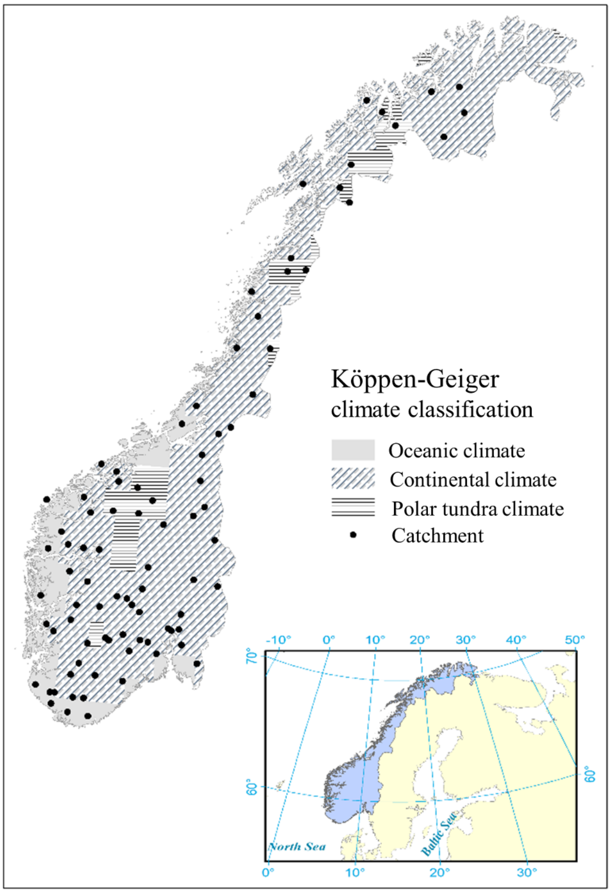


Fig. 1. The location of the study catchments and the modified Köppen-Geiger climate classification.

rainfall during autumn. The continental climate group displays low seasonality for precipitation but high seasonal variations in temperature, resulting in one peak runoff caused by snowmelt. The climate characteristics for the polar tundra climate group are similar to those of the continental group, but with lower temperature, and the snowmelt-induced peak in runoff occurs later.

Table 1 shows the average annual and seasonal precipitation, temperature and runoff for the three climate classes. Precipitation in the oceanic climate group is substantially larger than that in the other two groups, which show rather similar precipitation amounts. For temperature, the oceanic climate group shows the highest values, whereas the coldest temperatures are recorded in the polar tundra climate group. In particular, for the oceanic group, precipitation increases from the calibration to verification period for the winter season, but for the summer season, the difference is small between the two periods. For temperature, the increase from the calibration to verification period is smallest in the oceanic region compared to the other regions. The seasonal characteristics in runoff are similar to those of precipitation. Note that summer runoff decreases from the calibration to the verification period for all groups.

Since there is no potential evapotranspiration (Ep) data available in our study area, which are needed as the input data for the hydrological models, we applied the Hargreaves equation (Hargreaves, 1975) to calculate Ep (mm/day), which is recommended by Shuttleworth (1993) and Xu and Singh (2002):

$$E_p = 0.0023R_a(TC + 17.8)\sqrt{TR} \tag{1}$$

where $R_a$is the extraterrestrial radiation for the location in mm/day evaporation equivalent (Allen et al., 1998), TC is the temperature (°C), and TR is the daily temperature range (°C).

A set of catchment descriptors is needed for two of the regionalization methods, namely, the physical similarity and regression methods (see Table 2). These catchment descriptors were used in Yang et al. (2018, 2019). Similar catchment descriptors have been used in several studies for evaluating regionalization methods (e.g., He et al., 2011; McIntyre et al., 2005; Merz and Blöschl, 2004).

3. Methods

3.1. Hydrological models

Four widely used conceptual rainfall-runoff models running at a

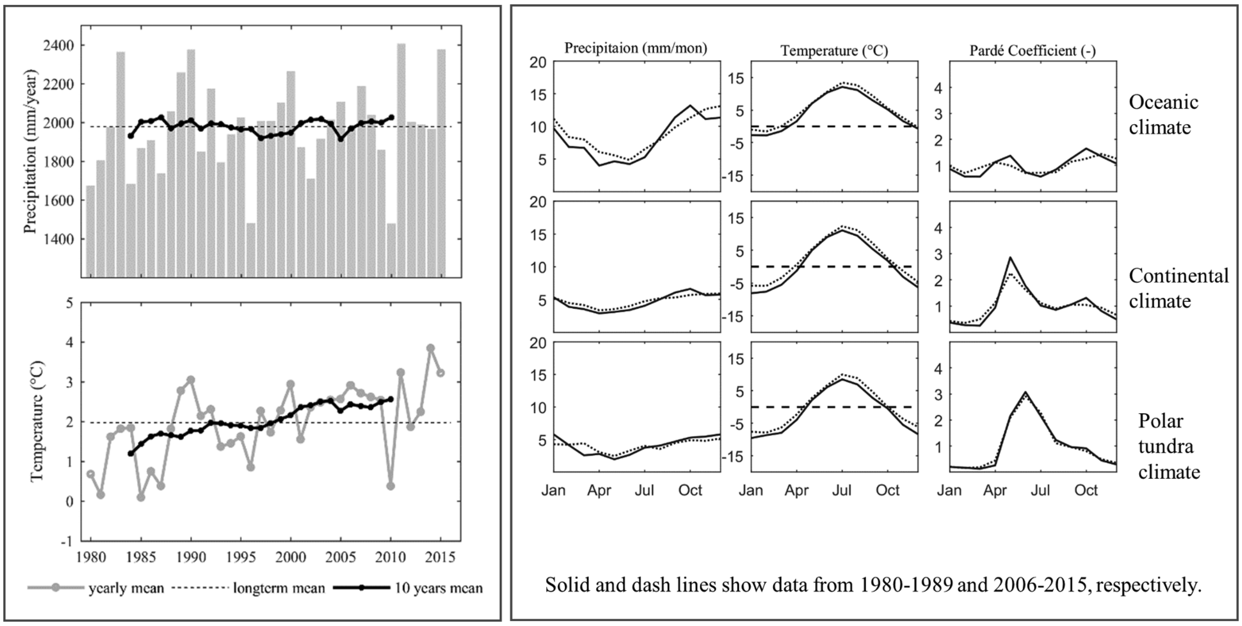


Fig. 2. The left panel shows yearly mean precipitation and temperature for the available data period, including a moving average with a sample window covering 10 years of data. The right panel shows the climatological distribution of precipitation, temperature and Pardé coefficient (i.e., ratio of the average monthly discharge to the mean annual runoff) using monthly data for the three climatic regions.

Table 1
The average precipitation, temperature and runoff information for all climate groups.

		Precipitation (mm/period)		Temperature (°C)		Runoff (mm/period)	
		calibration	validation	calibration	validation	calibration	validation
Oceanic climate	Year	2949	3211	4.1	5.2	2158	2342
	summer*	1411	1412	9.0	9.8	1197	1128
	winter	1508	1800	−0.7	0.5	961	1214
Continental climate	Year	1686	1750	0.8	2.3	1213	1250
	summer*	867	873	7.0	8.0	898	835
	winter	819	878	−5.3	−3.4	315	415
Polar tundra climate	Year	1633	1688	0.0	1.4	1187	1236
	summer*	817	819	6.1	7.1	942	908
	winter	816	869	−6.1	−4.3	245	328

*Summer is from 1st of May to 31st of October.

Table 2
The statistical information about catchment descriptors used in regionalization methods.

	Mean	Median	Minimum	Maximum
Area (km2)	340	145	3	5621
Climate index				
Mean annual precipitation (mm)	2255	1922	601	6008
Precipitation seasonality indices(1)	3.1	2.9	1.7	7.0
Mean annual temperature (°C)	2.7	2.5	−2.2	7.3
Temperature seasonality indices(2)	15.5	15.4	7.5	24.2
Aridity index(3)	0.1	0.1	0.0	0.4
Terrain characteristics				
Mean slope (°)	11	9	2	26
Mean elevation (m)	666	590	157	1472
Land use				
Artificial (%)	0.5	0.0	0.0	8.0
Agriculture (%)	4.1	1.1	0.0	57.6
Forest (%)	84.7	87.8	34.8	100.0
Wetland (%)	7.0	2.2	0.0	41.6
Waterbody (%)	3.7	2.9	0.0	15.1

(1) Precipitation seasonality indices: the ratio between the three consecutive wettest and driest months for each watershed.
(2) Temperature seasonality indices: the mean temperature of the hottest month minus the mean temperature of the coldest month in °C.
(3) Aridity index: the ratio between annual mean precipitation and potential evapotranspiration.

daily time step were selected for the analysis in this study, and a snow module was included in the models since runoff in many of the catchments is strongly affected by the accumulation and melting of snow. The number of model parameters varies from 6 to 17 between the models after adding the snow routine. Figure 3 shows the model structures, and a description of the parameters is available in Table 3.

GR4J (Génie Rural à 4 paramètres Journalier) is a model based on unit hydrograph principles with four free parameters (Perrin et al., 2003). It has been widely used in regionalization studies worldwide, such as in France (Oudin et al., 2008), China (Li et al., 2014) and Australia (Zhang et al., 2014, 2016). We coupled the GR4J model with a degree-day type snow module called CemaNeige that was developed by Valéry (2010). This snow module allows us to estimate snowmelt and simulate snowpack evolution using two additional parameters, and the coupling of GR4J and CemaNeige has been tested in other studies (e.g., Coron et al., 2014; Hublart et al., 2015).

WASMOD (The Water And Snow balance modelling system) is a model with simple structure and has been validated in many different climate regions (e.g., Xu and Singh, 2002; Li et al., 2013, 2015; Widén-Nilsson et al., 2007; Xu and Halldin, 1997). For regionalization studies, it has been applied in Sweden (Xu, 2003), Denmark (Muller-Wohlfeil et al., 2003) and Norway (Yang et al., 2018, 2019). The version of WASMOD used in this study has eight free parameters.

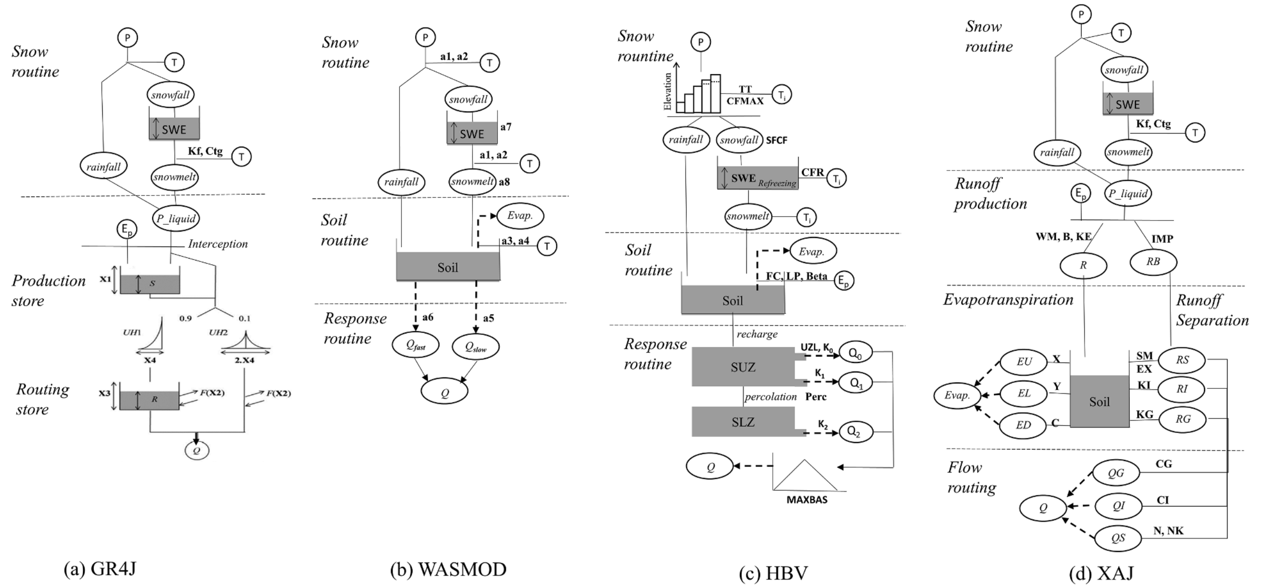
HBV (Hydrologiska Byråns Vattenbalansavdelning) is a popular model used for runoff simulation in both gauged and ungauged basins. For regionalization studies, it has been applied in different climate regions, such as Austria (e.g., Merz and Blöschl, 2004; Parajka et al., 2005), Sweden (Seibert and Beven, 2009), China (Jin et al., 2009), Canada (Samuel et al., 2011) and the US (Pool et al., 2017). In our study, we followed the structure and formulas in the HBV-light version (Seibert and Vis, 2012), which includes a snow routine, soil moisture routine, response function and routing routine. In total, this model has thirteen calibration parameters.

The XAJ (Xin An Jiang) model was developed for humid regions in China by Zhao et al. (1980, 1992) and has since become a widely used model in flood forecasting, water resources assessment, and climate change assessments. The original model consists of modules for computing evapotranspiration, runoff production, runoff separation, and flow routing. It has also been applied in many regionalization studies (e.g., Zhang and Chiew, 2009; Li et al., 2009, 2017). We implemented the structure shown in Lin et al. (2014) without the Muskingum routing module because our catchments are rather small in size with steep slopes, and therefore, river flow routing is not an important process (Li et al., 2014). However, there is no snow module in XAJ, and therefore, we coupled it with the CemaNeige snow module (see description of the GR4J model above). This model system contains seventeen parameters in total.

3.2. Regionalization methods

Spatial proximity, physical similarity and regression methods are commonly used in regionalization studies (e.g., Oudin et al., 2008; Petheram et al., 2012; Hrachowitz et al., 2013). For spatial proximity and physical similarity methods, which are classified as distance-based regionalization methods according to He et al. (2011), the model parameter values in ungauged catchments are transferred from gauged donor catchments. For the regression method, the model parameter values in ungauged catchments are determined by regression functions established using data from gauged basins. The regression method in this study is principal component regression (PCR), which couples principal component analysis (PCA) with the multiple linear regression method. Using PCA, a set of observations of possibly correlated catchment descriptors is converted into a set of linearly uncorrelated variables called principal components. Then, the relationships among model parameters and selected catchment descriptors are established using multiple linear regression. Finally, these functions are used for estimating model parameters in the ungauged catchments. Table 4 describes the equations and assumptions for the regionalization methods applied in this study.

For distance-based regionalization methods, i.e., spatial proximity and physical similarity, two approaches are often used for transferring the model parameters from the gauged donor to the ungauged target catchments (e.g., McIntyre et al., 2005; Oudin et al., 2008). (a) For the so-called parameter averaging option, the model parameters from the



(a) GR4J	(b) WASMOD	(c) HBV	(d) XAJ

P: Precipitation (mm/day)	T: Temperature (°C)	SWE: Snow Water Equivalent (mm)	E: Evapotranspiration (mm/day)	Q: Discharge (m3/s)

Fig. 3. The structure of hydrological models tested in this study. The circles show the input variables, the ellipses present the process/output variable and the model parameters are marked with bold text. For detailed model equations, please refer to the references for the (a) GR4J model (Perrin et al., 2003; Valéry, 2010), (b) WASMOD model (Xu, 2003), (c) HBV model (Seibert and Vis, 2012), and (d) XAJ model (Lin et al., 2014).

donor catchments are first averaged and then used to run the model for the target catchment. (b) For the so-called output averaging option, the model is first run using the parameter sets from the donor catchments (i.e., basins with runoff where model calibration is possible) on the target catchment and the outputs from the model are then averaged. As a result, there are five regionalization approaches used in this study, as shown in Table 5. For a more detailed description, please see Yang et al. (2018, 2019).

3.3. Performance evaluation

3.3.1. Model calibration and verification

In this study, we applied a widely used objective function proposed by Viney et al. (2009a) when calibrating the models. This objective function is a weighted combination of the Nash and Sutcliffe efficiency (Nash and Sutcliffe, 1970) and a logarithmic penalty function based on the bias as follows:

$$F = NSE - 5 * |\ln(1 + bias)|^{2.5} \tag{2}$$

where:

$$NSE = 1 - \frac{\sum (Q_{sim} - Q_{obs})^2}{\sum (Q_{obs} - \overline{Q_{obs}})^2} \tag{3}$$

$$bias = \frac{\overline{Q_{sim}} - \overline{Q_{obs}}}{\overline{Q_{obs}}} \tag{4}$$

Q_{obs} represents the observed runoff, and Q_{sim} represents the simulated runoff. F values can vary from $-\infty$ to the optimal value of 1. This objective function can come close to maximizing Nash and Sutcliffe efficiency (NSE) and minimizing the bias at the same time (Vaze et al., 2010). For the calibration process, we used a standard gradient-based automatic optimization method (Lagarias et al., 1998) implemented in the MATLAB software package ("fmincon" function; MATLAB R2016b, The MathWorks, Inc., Natick, Massachusetts, United States).

The split-sample test is commonly used for model verification, aiming to show the model validity in different climate conditions (e.g., Coron et al., 2012; Xu, 1999; Klemeš, 1986). In the current study, we evaluate the model performance for 1980–1989 and 2006–2015, and the temperature and precipitation in the latter period are approximately 1.4 °C and 5% higher than that in the first period.

3.3.2. Evaluation of regionalization methods

We performed three different evaluations of the regionalization methods. In the first evaluation, the performance of the regionalization methods was tested for all models using data from the calibration period, aiming to show the differences among the models. In this step, we applied a leave-one-out cross verification method as in many other studies (e.g., Yang et al., 2018; McIntyre et al., 2005). In the second analysis, we repeated the same evaluation but for the warmer and wetter verification period. This analysis thus tests the transferability of both the regionalization methods and hydrological models under climate change conditions (e.g., Broderick, 2016; Li et al., 2012). In the final evaluation, we summarize and discuss the performance of the regionalization methods for the three different climatic regions (see Section 2.1). Since the climate is changing to be warmer in the future (IPCC, 2014), the following regionalization performance for different climate conditions is investigated from 1980 to 1989 (calibration) to 2006–2015 (verification).

3.3.3. Evaluation criteria

To investigate the performance from different aspects, we applied four different criteria in this study. The calibration function F (Eq. (2)) is the first selection since it considers both the goodness of fit and the water balance aspects between the simulated and observed runoff. NSE (Eq. (3)) is the second evaluation criterion, which is the most commonly used criterion in hydrology to measure the fit of the hydrographs between the observed and simulated runoff, and is relatively sensitive to high flow (e.g., Oudin et al., 2008; Pushpalatha et al., 2012; Zhang and Chiew, 2009). Thus, we included another criterion, NSElog, which is based on the same formulation as NSE but computed on logarithmic transformed flows and with more emphasis on low flow (e.g., Oudin et al., 2008; Pushpalatha et al., 2012). Finally, the percentage of bias (Pbias) (Eq. (4)) is applied to measure the average tendency of the simulation to be larger or smaller than the observed counterparts.

The range for F, NSE and NSElog is $(-\infty, 1)$, where 1 means the simulated runoff perfectly fits the observed runoff and less than 0 suggests that the model is no better than the observed mean value. For Pbias, it varies between $(-\infty, +\infty)$ with the optimal value equal to 0

Table 3
Description of the calibrated model parameters in this study.

Parameter	Explanation	Reference
CemaNeige		Valéry (2010)
C_{TG}	Ponderation coefficient	
K_f	Degree-day factor	
GR4J		Perrin et al. (2003)
X1	Production store maximal capacity	
X2	Catchment water exchange coefficient	
X3	One-day maximal capacity of the routing reservoir	
X4	HU1 unit hydrograph time base	
WASMOD		Xu, (2003)
a1	Threshold temperature for rainfall and snowfall	
a2	Threshold temperature for snowpack and snowmelt	
a3	Proportion parameter in potential evapotranspiration	
a4	Exponent parameter in actual evapotranspiration	
a5	Proportion coefficient of base flow	
a6	Proportion coefficient of fast flow	
a7	Coefficient for snowpack	
a8	Coefficient for snowmelt	
HBV		Seibert and Vis, (2012)
TT	Threshold temperature	
CFMAX	Degree-day factor	
SFCF	Snowfall correction factor	
CFR	Refreezing coefficient	
FC	Field capacity	
LP	Threshold for reduction of evaporation	
Beta	Shape coefficient	
UZL	Threshold parameter for upper zone	
K0	Recession coefficient in upper zone	
K1	Recession coefficient in upper zone	
K2	Recession coefficient in lower zone	
Perc	Maximal flow from upper to lower box	
MAXBAS	Routing, length of weighting function	
XAJ		Lin et al. (2014)
WM	Areal soil moisture storage capacity	
B	The exponent of the soil moisture storage capacity curve	
KE	Ratio of potential evapotranspiration to pan evaporation	
IMP	Ratio of the impervious to the total area of the basin	
X	Proportion of soil moisture storage capacity of the upper layer to WM	
Y	Proportion of soil moisture storage capacity of the lower layer to WM	
C	Coefficient of deep evapotranspiration	
SM	Areal mean free water capacity of the surface soil layer	
EX	Exponent of the free water capacity curve	
KI	Coefficient of the free water storage to interflow	
KG	Coefficient of the free water storage to ground flow	
N	Number of reservoirs in the instantaneous unit hydrograph	
NK	Common storage coefficient in the instantaneous unit hydrograph	
CI	Recession constant of the lower interflow storage	
CG	Recession constant of the groundwater storage	

and worse performance for water balance simulation if the absolute Pbias is larger.

4. Results

4.1. Hydrological model performance in cross verification

Before evaluating both the hydrological models and the regionalization methods, we first assessed the performance of the models by a split-sample test. Figure 4 presents the cumulative density function (CDF) curves for all hydrological models over 86 catchments, measured by F value during 1980–1989 and 2006–2015.

For the first calibration period 1980–1989 (the left panel in Figure 4), the CDF curves from all the hydrological models stay close, and XAJ appears to be slightly better. The average F value is approximately 0.75 for XAJ, 0.73 for WASMOD, 0.72 for HBV and 0.69 for GR4J. In the verification period 2006–2015, the models perform differently, meaning the temporal transferability varies between the hydrological models. However, the best performance is still produced by XAJ, whose mean F value is approximately 0.68, followed by WASMOD (0.64). The HBV model shows the worst performance, with a mean F value of approximately 0.61 and the highest degradation of performance between the calibration and verification periods.

The results in the right panel (calibration in 2006–2015 and verification in 1980–1989) shows very similar characteristics to those in the left panel. XAJ produced the best performance for both the calibration and the verification periods. Following the rating classification from Moriasi et al. (2007), who labeled the performance as ‘good’ if NSE is larger than 0.65 and |Pbias| is less than 15%, the F values larger than 0.61 are considered “good” model performance. Considering the average aspect, all mean F values for our split-sample test are higher than 0.61. Thus, all hydrological models applied in the current study are classified as ‘good’ performing models for runoff simulation for both calibration and verification periods.

Table 6 gives the average model performance corresponding to the split-sample test by using other assessment criteria. First, regarding the water balance aspect, all models yield similarly ‘good’ performance for both subperiods with |Pbias| values smaller than 5%. Second, the model performance measured by NSE shows consistent findings with the results from the F value, i.e., (a) the models show similar performance in the calibration period but perform differently in the verification period; (b) XAJ is considered the best-performing model for both the calibration and the verification cases; and (c) HBV shows the largest decline in performance from the calibration to the verification period. This similarity between the results from the F value and NSE can be explained by the small Pbias for all the simulation results. Finally, according to the results of NSElog, which is more sensitive to low flow, the simple models (GR4J and WASMOD) display higher values in the calibration period, while WASMOD and XAJ show better performance in the verification period. Considering the performance loss from calibration to verification, relatively larger degradation appears for the NSElog than for the NSE and Pbias, especially for the GR4J model.

4.2. Evaluation of regionalization methods

4.2.1. Influence of the number of donor catchments on performance under stationary conditions

Figure 5 shows that the output averaging option gives better average performance than the parameter averaging option in both spatial proximity and physical similarity methods and for all the models, except for the case of one donor catchment, where both options provided the identical results as expected. When considering the number of donor catchments, the largest increase in performance typically occurs when changing from using one donor catchment to using two donor catchments, with the parameter option for XAJ as the only exception. This is in line with earlier studies that the number of donor catchments typically affects the performance of distance-based regionalization methods (e.g., Oudin et al., 2008; Yang et al., 2018). However, the number of donor catchments providing the best performance differs among the hydrological models and regionalization methods. For instance, for XAJ, two donor catchments give the best results for SP-out, whereas 8 donor catchments are needed for HBV to achieve the optimal performance. Finally, the difference in performance between the output and parameter averaging options increases with the

Table 4
Assumptions and descriptions of regionalization methods used in this study.

Method	Equation	Assumption and Description	Application examples
Spatial proximity	$D_{td} = \sqrt{(x_t - x_d)^2 + (y_t - y_d)^2}$	Closer basins show similar hydrological characteristics. The donor catchments are determined by the distanceD_{td}. x,y shows the location information, which uses the Universal Transverse Mercator (UTM) coordinate system.	Merz and Blöschl (2004), Oudin et al. (2008), Yang et al. (2018, 2019)
Physical similarity	$SI_{td} = \sum_{i=1}^{k} \frac{\lvert CD_{d,i} - CD_{t,i}\rvert}{\Delta CD_i}$	Similar attributes show similarly in terms of hydrological processes. The donor catchments are decided by the similarity index SI_{td}. CD is the catchment descriptor, shown in Table 2 in this study.	Burn and Boorman (1993), Poissant et al. (2017), Yang et al. (2018, 2019)
Regression	$MP_j = f_j(CD_i)$	A well-behaved relationship exists between the observable CDs and model parameters (MP), and the CDs used in regression provide information relevant to hydrological behavior at ungauged sites. The relationship (linear regression function), which is built on gauged basins, will be transferred to ungauged catchments.	Young (2006), Oudin et al. (2008), Merz et al. (2006), Yang et al. (2018, 2019)

t: target catchment.
d: donor catchment.
i: ith catchment descriptors.
k: total number of catchment descriptors.
j: jth model parameter.
CD: catchment descriptor. The climate indices in CDs varied from the calibration to verification period, others are assumed as constant.

Table 5
The tested regionalization methods in this study.

Regionalization methods	Abbreviation
Spatial proximity methods with parameter average option	SP-par
Spatial proximity methods with output average option	SP-out
Physical similarity methods with parameter average option	Phy-par
Physical similarity methods with output average option	Phy-out
Principal Component Regression method	PCR

Table 6
Average model performance in terms of Pbias, NSE and NSElog over the tested catchments in the split-sample test.

		calibration		verification	
		1980–1989	2006–2015	2006–2015	1980–1989
Pbias	GR4J	−0.81	−0.49	−4.37	−2.32
	WASMOD	2.61	3.15	−0.54	3.26
	HBV	−1.62	−1.49	−3.69	−3.90
	XAJ	−2.34	−1.69	−3.48	−1.80
NSE	GR4J	0.76	0.76	0.67	0.66
	WASMOD	0.77	0.76	0.68	0.67
	HBV	0.77	0.76	0.65	0.65
	XAJ	0.79	0.78	0.72	0.71
NSElog	GR4J	0.74	0.75	0.39	0.37
	WASMOD	0.67	0.71	0.58	0.55
	HBV	0.37	0.51	0.28	0.33
	XAJ	0.51	0.65	0.52	0.55

number of model parameters. For example, the difference in the average F value between the two options for the GR4J model was approximately 0.025 and increased to 0.075 for XAJ. Thus, when using a model with many parameters, it is more important to use the output averaging option to achieve optimal performance for runoff simulations in ungauged basins.

The physical similarity methods require fewer donor catchments to achieve optimal performance for runoff simulations in ungauged basins compared to that for the spatial proximity methods (Table 7). On average, the best performance by the physical similarity methods was produced by 3 donor catchments, whereas the corresponding number for the spatial proximity methods was 8. It is also noteworthy that the parameter averaging option requires fewer donor catchments than the output averaging option for both the physical similarity and the spatial

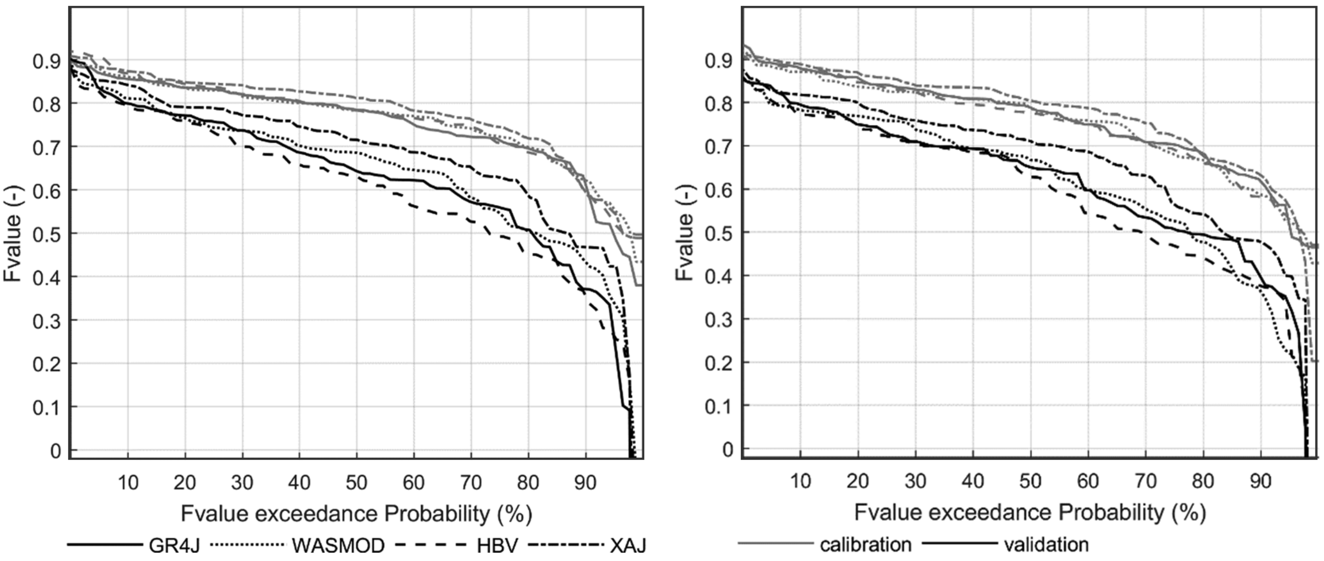


Fig. 4. The performance of hydrological models by split-sample test evaluated by the F value over 86 catchments. The left panel shows the results for calibration in 1980–1989 and verification in 2006–2015; the right panel displays the results of calibration in 2006–2015 and verification in 1980–1989.

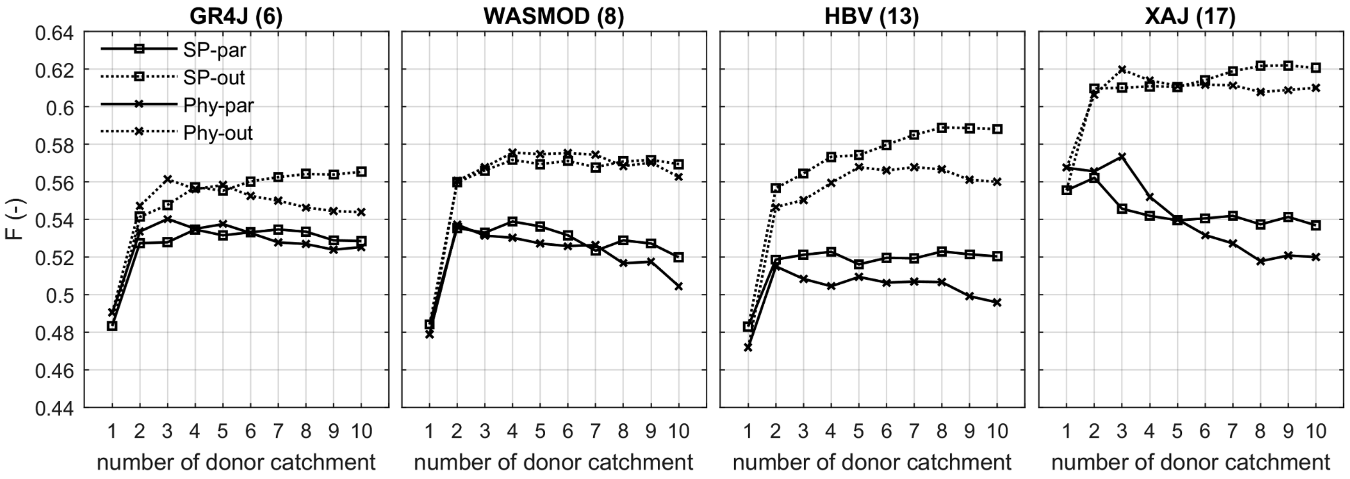


Fig. 5. Model performance versus number of donor catchments for the distance-based regionalization methods and four different models. The number of model parameters is given in the parenthesis next to the model name.

proximity methods. Therefore, for practical applications, it is highly recommended to analyze the relationship between the regionalization performance and the number of donor catchments to choose the best configuration to obtain the optimal results for each case.

4.2.2. *Regionalization performance assessment for all catchments*

As discussed in Section 2.2 (Figure 2 and Table 1), the climate conditions, especially air temperature, differed between 1980 and 1989 and 2006–2015. This section presents the influence of climate conditions on regionalization performance when the models are calibrated in 1980–1989. The evaluation results presented here applied the optimized number of donor catchments for each method and model, as shown in Table 7.

4.2.2.1. Comparison of regionalization performance between hydrological models. Figure 6 shows the distribution of F values as split violin plots for the five regionalization methods and four hydrological models for both the calibration and verification periods. Foremost, for all the hydrological models, the regionalization methods applying the output averaging option (SP-out and Phy-out) showed better performance than the parameter averaging option (SP-par and Phy-par), and the regression method is the worst (compare black dots with circles). This ranking applies for both the calibration and the verification periods, where the methods with output averaging options presented more negative skewed distributions and higher mode values than those of the other methods. On the other hand, for both periods, the difference in the average performance between the regionalization methods is smaller for GR4J than for the other models. This difference seems to increase with the number of model parameters and is thus largest for XAJ. For instance, in the calibration period, the range in the average F values between the regionalization methods equals 0.04 for GR4J and 0.09 for XAJ. Finally, from the calibration to verification period, performances decreased for all the hydrological models and regionalization methods but to various extents. Measured by the decrease in the overall mean F values from the calibration (solid line) to verification (dashed line) period, HBV and XAJ displayed larger declines in performance than those of GR4J and WASMOD.

Table 7
The number of donor catchments providing the best performance for each regionalization method and hydrological model in the leave-one-out cross validation.

	GR4J	WASMOD	HBV	XAJ
SP-par	7	4	8	2
SP-out	10	9	8	9
Phy-par	3	2	2	3
Phy-out	3	5	5	3

Figure7 compares the regionalization performance in terms of the average values of Pbias, NSE and NSElog for all catchments using four hydrological models in the calibration and verification periods. Appendix A presents the violin plot for the evaluation criteria over all the tested catchments.

Regarding the water balance simulation, all average values of Pbias vary within (−10%, 10%). The smallest water balance error for regionalized runoff simulation varies with the hydrological models and regionalization methods. In general, SP-out and Phy-out tend to yield smaller errors for water balance simulation than those of the other methods.

The NSE results give similar findings as the F value. First, SP-out and Phy-out methods perform best for all the hydrological models, with all average NSE values larger than 0.6, and PCR performs worst. Second, the difference in NSE between the regionalization methods increases with the growing number of parameters for the hydrological models. For example, the regionalization performance in the calibration period ranges within (0.57, 0.61) for GR4J and (0.57, 0.67) for XAJ. Third, relatively larger degradation of the average regionalization performance is found using the HBV and XAJ models from the calibration to the verification period.

For the low-flow evaluation, the regionalization methods with the output average option (SP-out and Phy-out) substantially outperform the other methods, and the performance differences between the regionalization methods are more distinct for HBV and XAJ. Furthermore, the average performance of the regionalization methods is highly influenced by the hydrological models. In this study, WASMOD and HBV produced the highest and lowest average NSElog values for the regionalization methods, respectively. Compared with the results from the NSE and F values, the evaluation by NSElog presents a more recognizable performance difference between the regionalization methods and hydrological models, as well as the difference between the two subperiods.

4.2.2.2. Comparison of performance between regionalization methods. Figure 8 compares the performance difference in terms of NSE and NSElog between the hydrological models for each regionalization method during the calibration and verification periods. We omit the results of the F value and Pbias in the following analysis due to high similarity between the results from the F value and NSE (see Figure 6 and Appendix A) and small average |Pbias| values (see Figure 7).

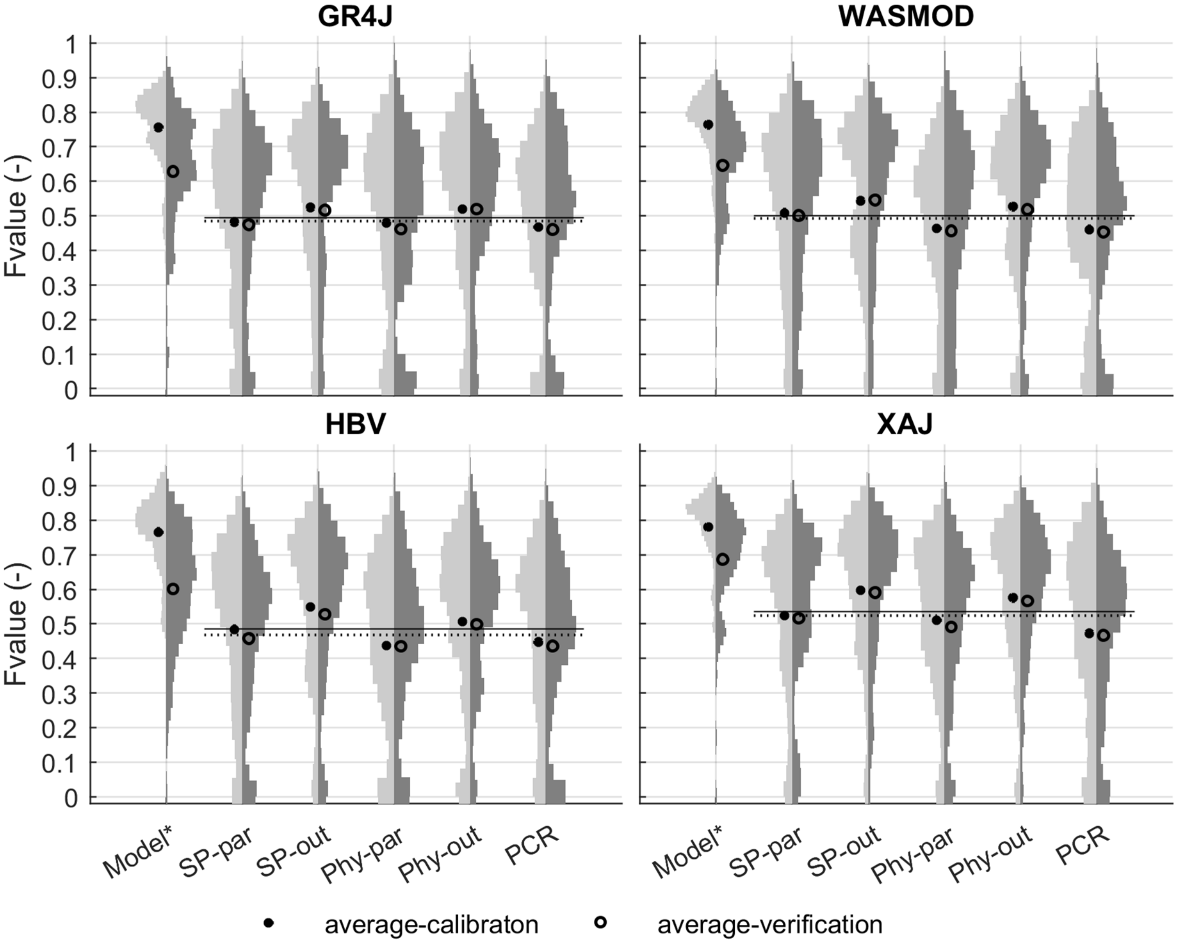


Fig. 6. Split violin plots show the distributions of F values for the five regionalization methods by each hydrological model during the calibration (left side of the violin) and verification (right side of the voilin) periods. For each model and regionalization method, the solid black dots show the average performance for the calibration period, whereas the black circle shows the corresponding value for the verification period. The average performance of all regionalization methods for each hydrological model is shown as a solid line for the calibration period and as a dashed line for the verification period. The plot displays results from the 86 study catchments. The 'model' in the x-axis label shows the hydrological model performance in the calibration (left side of the violin) and validation (right side of the voilin) periods.

According to the average NSE values, XAJ is considered the best hydrological model for all the distance-based regionalization methods and the second best model for PCR. GR4J shows the best results for PCR, but the difference in performance between the models (the gray bars for PCR) is smallest among the regionalization methods, indicating that the hydrological models have relatively smaller influence on the regression method than on the distance-based methods. However, this difference is enhanced from the calibration to the verification period, indicating a larger influence of the hydrological model on future runoff predictions. According to NSElog, WASMOD shows the best performance for all the regionalization methods and for both periods. In general, a larger difference between the hydrological models appears for low flows (indicated by NSElog) than for high flows (indicated by NSE).

4.2.3. Assessment of regionalization performance for different climatic regions

The three climatic regions shown in Figure 1 display very different runoff regimes, particularly between the oceanic and the two remaining groups (Figure 2). For illustration purposes, the dependence of the performance of the regionalization methods on the geographical regions as measured by NSE is shown in Figure 9. It is seen that the oceanic region presented generally better regionalization performance than that of the other two regions, whose performance variation was smaller as well (only four performance classes shown on the figure). Then, some common characteristics are presented in all the regions. First, when considering the regionalization methods, the output averaging option tended to give higher performance than all the other methods. When focusing on the hydrological models, XAJ showed the best performance in most cases for both the calibration and verification periods. Otherwise, none of the remaining models consistently showed better results than the other models for all climatic regions and regionalization methods. Finally, GR4J produced the lowest variation in performance within the climatic regions between the regionalization methods in almost all cases. From the calibration to verification period, the highest ranking for XAJ with SP-out and Phy-out methods did not change.

5. Discussion

5.1. Hydrological model performance

According to the performance classification presented by Moriasi et al. (2007), the split-sample test result in our study indicated that all the hydrological models were able to provide 'good' simulations of runoff for both the calibration and the verification periods. Especially for the water balance simulation, the mean values of |Pbias| for all the studied models are smaller than 5%.

According to the evaluations in the calibration period based on the F value and NSE in our study area, XAJ is the best-performing model, and the performance tends to decrease with a decrease in the number of parameters for the hydrological models. This finding is in line with the

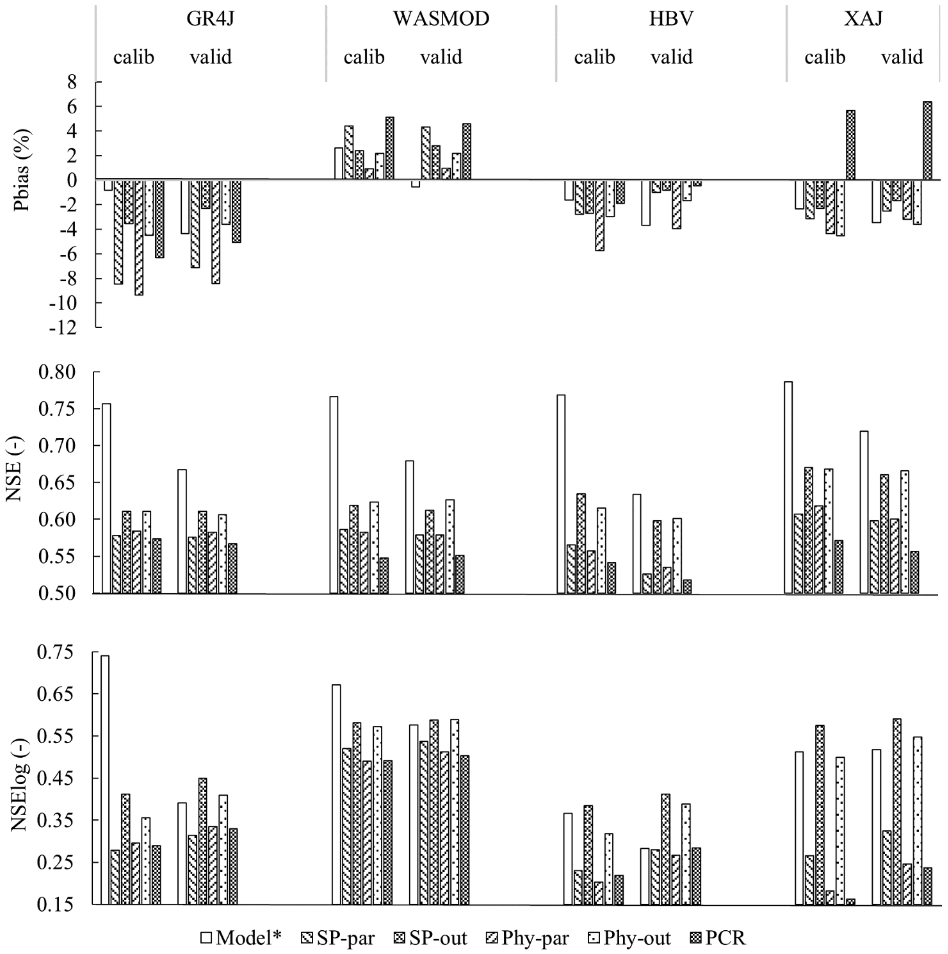


Fig. 7. Average performance for the different hydrological models and regionalization methods, given by Pbias, NSE and NSElog. Model* is the result of model simulation performance in the calibration ('calib') and verification ('valid') periods.

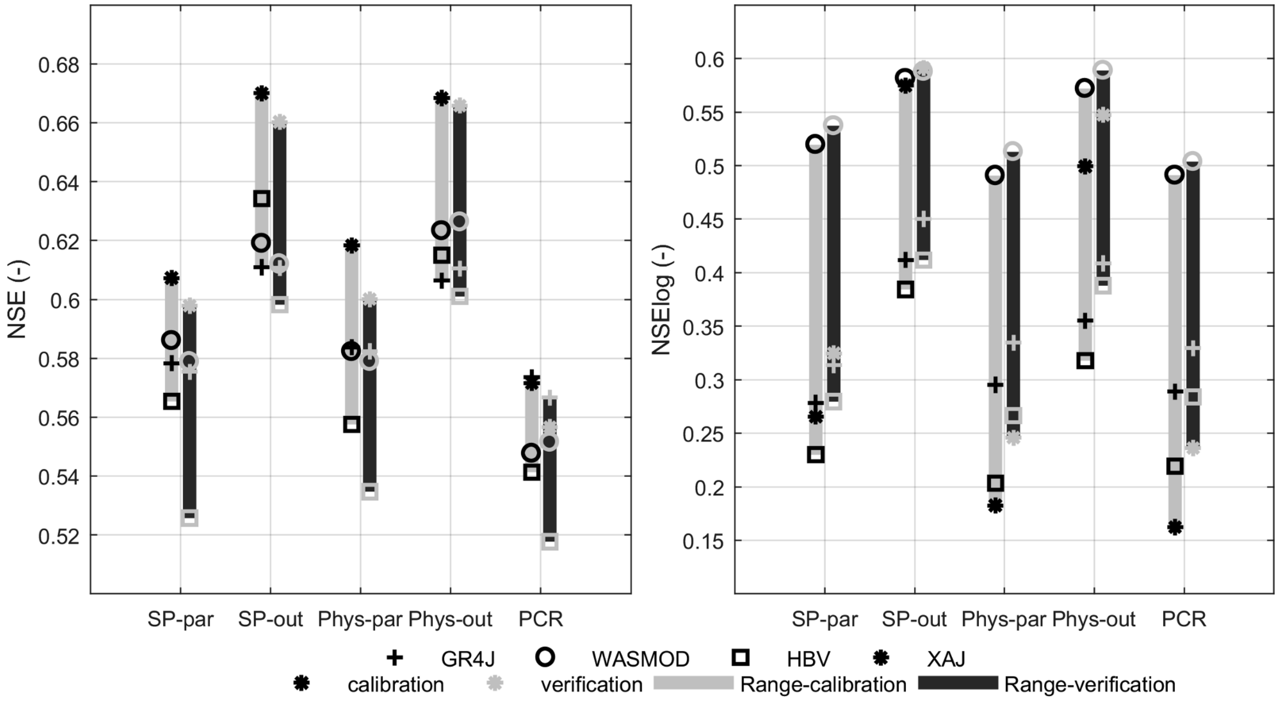


Fig. 8. Comparison of hydrological model performance over five regionalization methods in the calibration and verification periods. The bar shows the maximum difference between the hydrological models evaluated by the average NSE and NSElog values over 86 catchments.

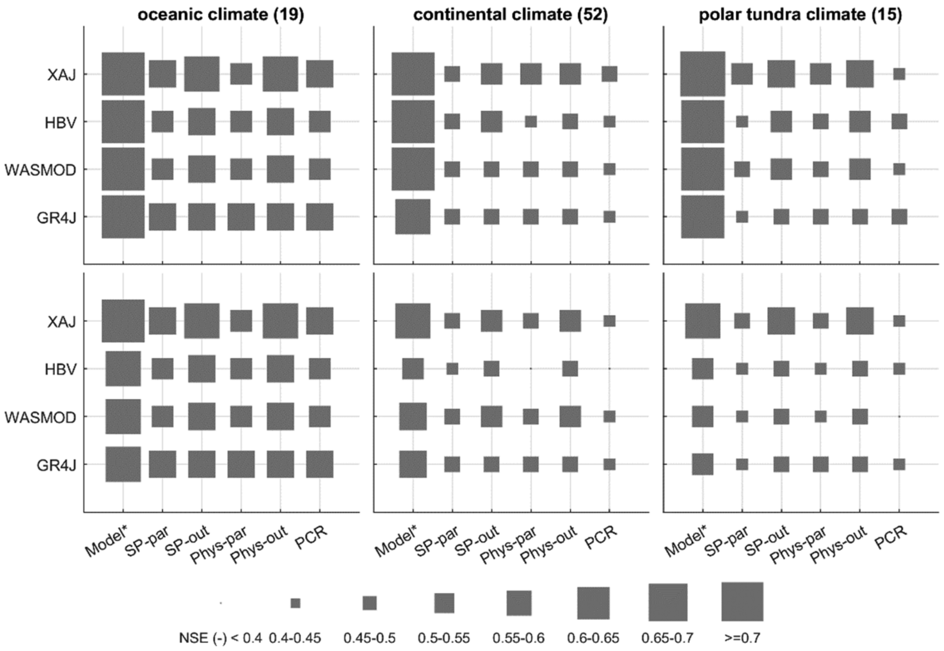


Fig. 9. The performance of the regionalization methods and hydrological models in different climatic regions. The size of the boxes is proportional to the average NSE value of the catchments within each climate group. The upper panel shows the results from the calibration period, and the lower panel shows the verification period. The number of catchments in each group is given in the title of each column. The 'Model' in the x-axis label shows the hydrological model performance for runoff simulation without regionalization.

statement that increasing the number of model parameters can lead to better performance during the calibration period (e.g., Perrin et al., 2001; Petheram et al., 2012; Parajka et al., 2013). However, the result in terms of low flow simulation (evaluations by NSElog) did not support that statement. For example, WASMOD outperformed XAJ and HBV for both subperiods. Therefore, further study is needed to assess the relationship between hydrological model complexity and performance in terms of low flow. Furthermore, for the verification results, the performances among the models varied substantially. The degradation of performance is quite similar between the hydrological models evaluating by the F value and NSE, but distinct differences are shown in the NSElog results. It reminds us that specific criteria are needed for evaluation of hydrological models when the emphasis stands on low flow or draughts. Regarding the model performance change from the calibration to the verification period, the model performance of the XAJ model did not vary substantially. This is incompatible with earlier findings, which suggest that a complex model tends to have less stable performance than simple models in the verification period (e.g., Perrin et al., 2001; Holländer et al., 2009). This phenomenon might relate to the model structure; for instance, the runoff concentration in the XAJ model includes surface runoff, interflow runoff and groundwater runoff with three parameters that may better represent the processes in our study catchments.

5.2. Evaluation of regionalization methods

5.2.1. Influence of the number of donor catchments on performance

To test the influence of the number of donor catchments on model performance, we examined the relationship between regionalization performance and the number of donor catchments for all the models with distance-based methods. The results indicate that using one donor catchment, which might be either the spatially nearest or physically most similar watershed, gives worse results than using a set of donor catchments. This conclusion is supported by all the tested models in our study, which is in line with previous findings (e.g., Arsenault and Brissette, 2014; Oudin et al., 2008). Multiple donor catchments typically provide more information than single donor catchments, which may explain the behavior described above (e.g., Viney et al., 2009b). However, the output averaging option might tend to smooth the flow variability as the number of donor catchments increases. This is especially the case if the donors give models with different time lags between rainfall and peak flow. Therefore, the smoothing effect and tradeoff between the benefits of gains in performance with "more information" and loss of performance due to this possible smoothing is worth further investigation in future studies. Our results additionally confirmed that the output averaging option provided better performance than the parameter averaging option in all the model and method combinations (e.g., Oudin et al., 2008, Bao et al., 2012; Yang et al., 2018). Since we applied hydrological models with different complexities and number of parameters, a promising and new finding is presented in this study: the difference in performance between the parameter averaging and output averaging options increases with the number of model parameters (see Figure 5). First, this result can be explained by the 'nonlinear independence' influence between model parameters; thus, transferring the linearly interpolated individual model parameter value (the parameter averaging option) will lead to unreasonable model parameters and results (Bárdossy, 2007). Second, hydrological models with more parameters tend to increase the interaction between their parameters (e.g., Perrin et al., 2003; Poissant et al., 2017). Hence, we should consider the model parameters as a whole set rather than individual values for regionalization research as suggested by Bárdossy (2007) and Oudin et al. (2008).

Some previous studies used one donor catchment for regionalization evaluation according to spatial or physical similarity and concluded that the difference in performance between hydrological models is negligible (e.g., Viney et al., 2009b; Chiew, 2010; Petheram et al., 2012). However, in the current study, XAJ produced distinct results from the other models (see Figure 5 results with 1 donor catchment), which suggests that the performance of regionalization methods is affected by the choice of hydrological models even with one donor catchment.

5.2.2. Assessment over hydrological models

Although we claimed that the methods with the output averaging option (SP-out and Phy-out) produced better performance than the other methods, it is difficult to determine the most appropriate method between the spatial proximity (SP-out) and physical similarity (Phy-out) methods (also valid for excluding the influence on the hydrological model performance of calibration and verification, see Appendix B). This is consistent with the evaluation by using one hydrological model (monthly WASMOD) in the same area by Yang et al. (2018). According to the explanation from Oudin et al. (2008), it is not possible to decide which approach (SP-out or Phy-out) is the most appropriate one when the streaming network density is lower than 60 stations per 100,000 km^2. As we used four hydrological models at different complexity levels, this result additionally confirmed that this assertion is independent of the selection of hydrological models.

Investigating the model preference for regionalization methods from different aspects, XAJ should be preferred when the evaluation is more focused on high flow, while WASMOD should be considered for low-flow analysis. This result is consistent with the model performance for gauged catchments (see Figure 4 and Table 6). This result tends to support the claim that there is no incentive to prefer a parsimonious hydrological model for regionalization studies rather than a model with adequate complexity (Arsenault et al., 2015; Poissant et al., 2017). However, hydrological models with fewer parameters are recommended when no preknowledge about the regionalization performance is available since the performance difference between the regionalization methods is relatively smaller. For the regression method, the model with more parameters works worse, probably due to the stronger interaction influence when increasing the number of parameters (e.g., Perrin et al., 2003; Poissant et al., 2017). Another limitation of the regression method is that not all the functions for the model parameters follow the linear assumption (e.g., Blöschl, 2005) and poor performance results from the accumulated errors.

5.2.3. Assessment in different climatic regions

According to both the NSE and NSElog results, SP-out and Phy-out perform best for all the climatic regions. Therefore, it seems reasonable to conclude that the selection of the climatic region has no large effect on the ranking of regionalization methods. However, the average regionalization performance in the oceanic climate region is substantially better and varies within a smaller range than in the other two cold regions. This indicates that the uncertainty in the selection of regionalization methods is larger in cold and dry regions than in warm and wet regions (see Figure 2). Due to the limited number of catchments in the oceanic climate and polar tundra climate regions, further comprehensive studies are needed to conclude the preferences of hydrological models and regionalization methods over various regions.

6. Conclusions

The main aim of this study was to investigate how different combinations of regionalization methods, hydrological models and climate conditions will influence the overall performance of hydrological simulations in ungauged basins. We assessed the performance of four hydrological models and five regionalization schemes (a) under stationary climate conditions to test how the performance of the regionalization methods depends on the choice of hydrological models, (b) under different climate conditions to assess the stability in performance of the hydrological models and regionalization methods as climate changes, and (c) in different climatic regions to test how the performances of the simulations vary between these regions. The study was performed using data from 86 catchments in Norway, covering three climatic groups according to the Köppen-Geiger classification.

In this study, we found that for all the hydrological models, the distance-based approaches with the output averaging option (SP-out and Phy-out) always outperformed the other tested methods, especially for the low-flow estimation. Second, the difference in performance between the output and parameter averaging options is not stable and positively increases with the number of parameters for the hydrological models. From our study, the performance difference between these options is the largest for XAJ and the smallest for GR4J. Third, the performance difference among the regionalization methods was smaller for models with fewer parameters (GR4J and WASMOD) compared to that of the models with more tunable parameters (HBV and XAJ). Regarding the model influence on regionalization performance, XAJ is recommended as the best-performing model according to the evaluations by NSE and F values, whereas NSElog recommends WASMOD as the best through the evaluation. Furthermore, clear differences in general were displayed for three climatic regions, and oceanic climatic regions provided the best performance and smallest variance over the regionalization methods and hydrological models. Moreover, the difference in hydrological model performance seems smaller among the regionalization methods than among the climate regions. From calibration to verification periods, the general performance for the regionalization methods did not show large degradations.

Although this study produced some solid conclusions that were not available before, there are some limitations of the current study. Compared with the general evaluation of hydrograph fit and water balance, assessment with emphasis on low flow showed more contrasting results, which requires closer attention in future work. In addition, studies with more different hydrological models are needed to show the influence of hydrological model selection on regionalization performance. Moreover, studies with more contrast in climate conditions are recommended to investigate the transferability of conclusions across climate regions and climate changing conditions, which is essential for future prediction.

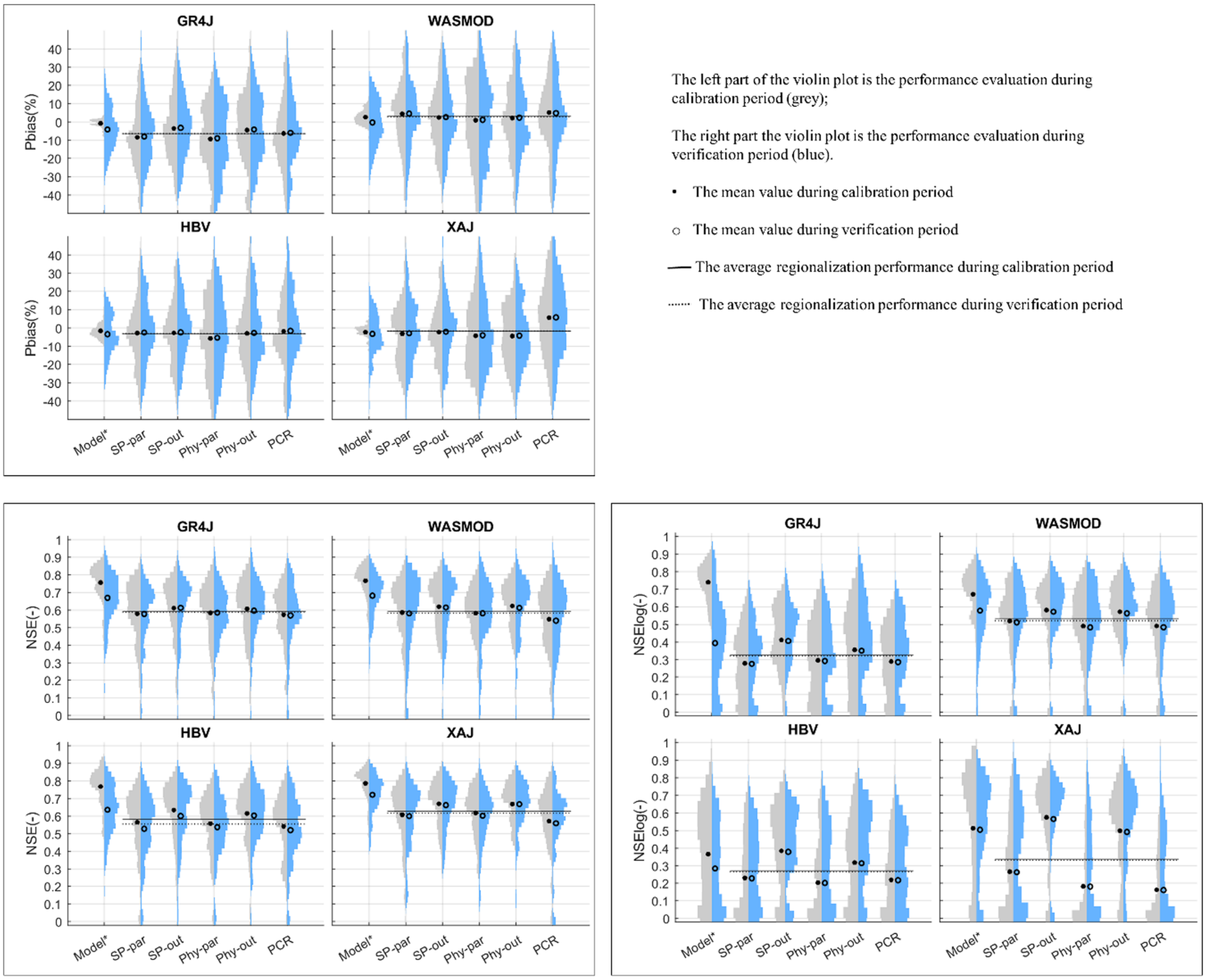
Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is supported by the Research Council of Norway (FRINATEK Project 274310), Research and Development Funding (Project number 80203) of the Norwegian Water Resources and Energy Directorate (NVE), and the China Scholarship Council. We would like to thank the NVE for providing the data for this study. We are thankful to the reviewers whose insightful and constructive comments have led to a significant improvement in the quality of the paper.

Appendix



The left part of the violin plot is the performance evaluation during calibration period (grey);

The right part the violin plot is the performance evaluation during verification period (blue).

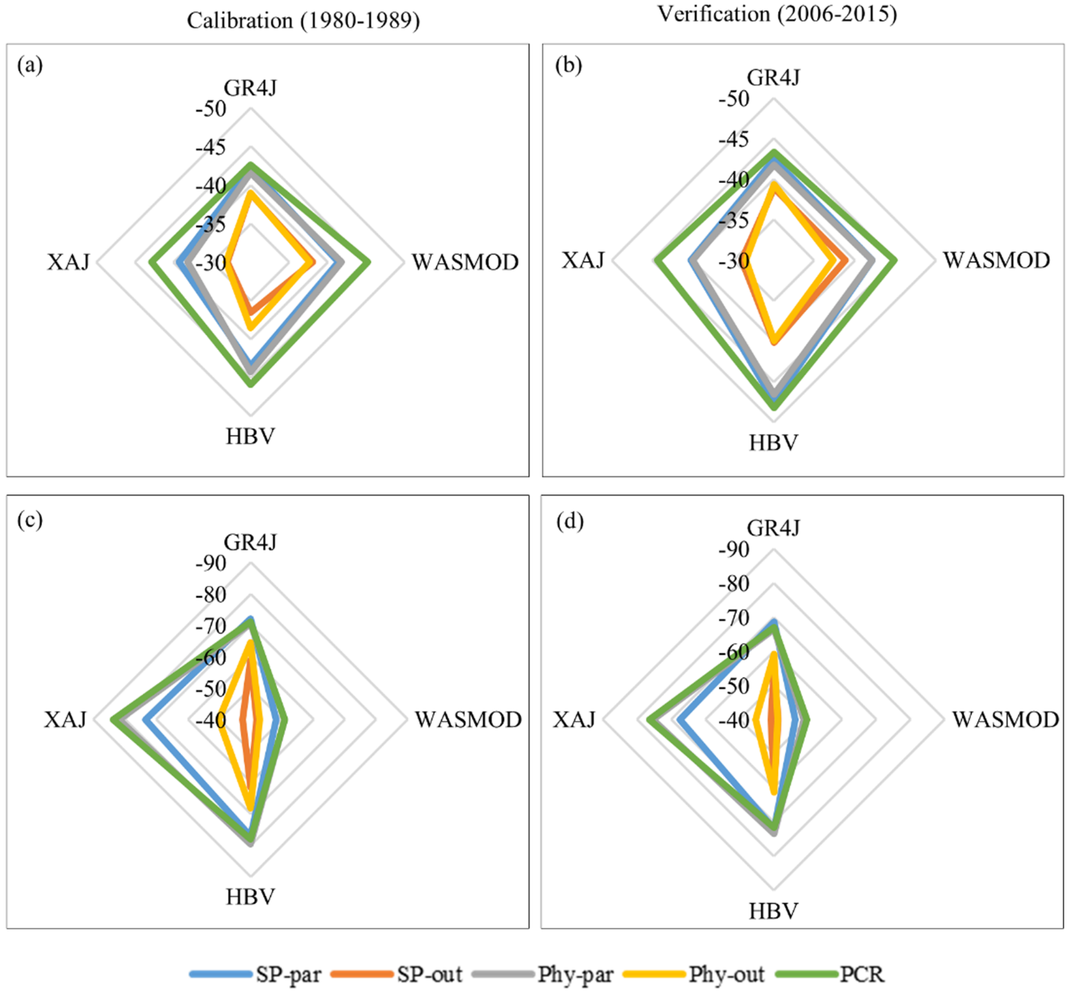
- The mean value during calibration period

○ The mean value during verification period

—— The average regionalization performance during calibration period

······ The average regionalization performance during verification period

Appendix A. The performance assessment over all tested catchments by Pbias, NSE and NSElog during calibration (left side of the violin) and verification (right side of the violin) periods. For each model and regionalization method, the solid black dots show the average performance for the calibration period, whereas the black circles show the corresponding values for the verification period. The average performance of all the regionalization methods for each hydrological model is shown as a solid line for the calibration period and as a dashed line for the verification period. Model* shows the hydrological model performance for calibration (left side of the violin) and verification (right side of the violin).



Appendix B. The performance evaluation by the relative average values of NSE (a and b) and NSElog (c and d) in the calibration and verification periods. The average relative NSE value for the kth hydrological model with the jth regionalization method is calculated as: $\Delta\overline{NSE}_{j,k} = \frac{1}{86} * \sum_{i=1}^{86} \frac{(NSE_{i,regn_j} - NSE_{i,model_k})}{NSE_{i,model_k}}$, where $NSE_{i,regn_j}$stands for the NSE value for jth regionalization method in i th tested basin and $NSE_{i,model_k}$ stands for the NSE value for hydrological model performance in the ith tested basin. Therefore, smaller values indicate the performance of the regionalization methods that is more similar to the performance of hydrological models. In this study, parameter k ranging in (1, 4) and j ranging in (1, 5) show the number of hydrological models and regionalization methods; 86 shows the number of tested basins. The same formula is used to calculate the average relative NSElog values.

References

Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998. Crop evapotranspiration, guidelines for computing crop water requirements, Irrig. and Drain. Pap. 56. U.N. Food and Agric. Organ., Rome.

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.

Arsenault, R., Poissant, D., Brissette, F., 2015. Parameter dimensionality reduction of a conceptual model for streamflow prediction in Canadian, snowmelt dominated ungauged basins. Adv. Water Resour. 85, 27–44. https://doi.org/10.1016/j.advwatres.2015.08.014.

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

Bárdossy, A., 2007. Calibration of hydrological model parameters for ungauged catchments. Hydrol. Earth Syst. Sci. Discuss. 3, 1105–1124. https://doi.org/10.5194/hessd-3-1105-2006.

Beck, H.E., Zimmermann, N.E., McVicar, T.R., Vergopolan, N., Berg, A., Wood, E.F., 2018. Present and future köppen-geiger climate classification maps at 1-km resolution. Sci. Data 5, 1–12. https://doi.org/10.1038/sdata.2018.214.

Blöschl, G., 2005. Rainfall–runoff modelling of ungauged catchments. In: Anderson, M.G. (Ed.), Encyclopedia of Hydrological Sciences. John Wiley & Sons, Chichester, pp. 2061–2080.

Blöschl, G., Montanari, A., 2010. Climate change impacts-throwing the dice? Hydrol. Process 24, 374–381. https://doi.org/10.1002/hyp.7574.

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.

Broderick, C., 2016. Transferability of hydrological models and ensemble averaging methods between contrasting climatic periods Ciaran. Water Resour. Res. Res. 8343–8373. https://doi.org/10.1002/2016WR018850.

Chiew, F.H.S., 2010. Lumped conceptual rainfall-runoff models and simple water balance methods: Overview and applications in ungauged and data limited regions. Geogr. Compass 4, 206–225. https://doi.org/10.1111/j.1749-8198.2009.00318.x.

Coron, L., Andre, V., Perrin, C., Lerat, J., Vaze, J., Bourqui, M., Hendrickx, F., 2012. Crash testing hydrological models in contrasted climate conditions : An experiment on 216 Australian catchments 48, 1–17. https://doi.org/10.1029/2011WR011721.

Coron, L., Andréassian, V., Perrin, C., Bourqui, M., Hendrickx, F., 2014. On the lack of robustness of hydrologic models regarding water balance simulation: a diagnostic approach applied to three models of increasing complexity on 20 mountainous catchments. Hydrol. Earth Syst. Sci. 18, 727–746. https://doi.org/10.5194/hess-18-727-2014.

Egbuniwe, N., Todd, D.K., 1976. Application of the stanford watershed model to nigerian watersheds. JAWRA J. Am. Water Resour. Assoc. 12, 449–460. https://doi.org/10.1111/j.1752-1688.1976.tb02710.x.

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

Hargreaves, G.H., 1975. Moisture availability and crop production. Trans. ASAE 18, 980–984.

Holländer, H.M., Blume, T., Bormann, H., Buytaert, W., Chirico, G.B., Exbrayat, J.F., Gustafsson, D., Hölzel, H., Kraft, P., Stamm, C., Stoll, S., Blöschl, G., Flühler, H., 2009. Comparative predictions of discharge from an artificial catchment (Chicken Creek) using sparse data. Hydrol. Earth Syst. Sci. 13, 2069–2094. https://doi.org/10.5194/hess-13-2069-2009.

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

Hublart, P., Ruelland, D., De Cortázar, GarcÍa, Atauri, I., Ibacache, A., 2015. Reliability of a conceptual hydrological model in a semi-arid Andean catchment facing water-use changes. In: IAHS-AISH Proceedings and Reports, pp. 203–209. https://doi.org/10.5194/piahs-371-203-2015.

IPCC, 2014. Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. In: Core Writing Team, R.K. Pachauri and L.A. Meyer (Eds.). IPCC, Geneva, Switzerland, 151 pp.

Jansson, A, Tveito, O E, Pirinen, P, & Scharling, M., 2007. NORDGRID - a preliminary investigation on the potential for creation of a joint Nordic gridded climate dataset. met.no Report 03/2007.

Jin, X., Xu, C., yu, Zhang, Q., Chen, Y.D., 2009. Regionalization study of a conceptual hydrological model in Dongjiang basin, south China. Quat. Int. 208, 129–137. https://doi.org/10.1016/j.quaint.2008.08.006.

Klemeš, V., 1986. Operational testing of hydrological simulation models. Hydrol. Sci. J. ISSN 6667. https://doi.org/10.1080/02626668609491024.

Kottek, M., Grieser, J., Beck, C., Rudolf, B., Rubel, F., 2006. World Map of the Köppen-Geiger climate classification updated. Meteorologishe Zeitschrift 15, 259–263.

Lagarias, J.C., Reeds, J.A., Wright, M.H., Wright, P.E., 1998. Convergence Properties of the Nelder-Mead Simplex Method in Low Dimensions. SIAM J. Optim. 9, 112–147. https://doi.org/10.1137/S1052623496303470.

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 1239–1254. https://doi.org/10.5194/hess-16-1239-2012.

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. https://doi.org/10.1016/j.jhydrol.2014.01.014.

Li, H., 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. https://doi.org/10.1016/j.jhydrol.2017.05.015.

Li, H., Zhang, Y., Chiew, F.H.S., Xu, S., 2009. Predicting runoff in ungauged catchments by using Xinanjiang model with MODIS leaf area index. J. Hydrol. 370, 155–162. https://doi.org/10.1016/j.jhydrol.2009.03.003.

Li, L., Diallo, I., Xu, C.Y., Stordal, F., 2015. Hydrological projections under climate change in the near future by RegCM4 in Southern Africa using a large-scale hydrological model. J. Hydrol. 528, 1–16. https://doi.org/10.1016/j.jhydrol.2015.05.028.

Li, L., Ngongondo, C.S., Xu, C.-Y., Gong, L., 2013. Comparison of the global TRMM and WFD precipitation datasets in driving a large-scale hydrological model in southern Africa. Hydrol. Res. 44, 770. https://doi.org/10.2166/nh.2012.175.

Lin, K., Liu, P., He, Y., Guo, S., 2014. Multi-site evaluation to reduce parameter uncertainty in a conceptual hydrological modeling within the GLUE framework. J. Hydroinform. 16 (1), 60–73. https://doi.org/10.2166/hydro.2013.204.

Magette, W.L., Shanholtz, V.O., Carr, J.C., 1976. Estimating selected parameters for the Kentucky Watershed Model from watershed characteristics. Water Resour. Res. 12, 472–476. https://doi.org/10.1029/WR012i003p00472.

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. Regionalization of catchment model parameters. J. Hydrol. 287, 95–123. https://doi.org/10.1016/j.jhydrol.2003.09.028.

Merz, R., Blöschl, G., Parajka, J., 2006. Spatio-temporal variability of event runoff coefficients. J. Hydrol. 331, 591–604.

Mohr, M., 2009. Comparison of Version 1.1 and 1.0 of gridded temperature and precipitation data for Norway. met.no Note 19/2009.

Moriasi, D.N., Arnold, J.G., VanLiew, M.W., Bingner, R.L., Harmel, R.D., Veith, T.L., 2007. Evaluation guidelines for systematic quantification of accuracy in watershed simulations. Trans. ASABE 50 (3), 885–900.

Muller-Wohlfeil, D.-I., Xu, C.-Y., Lversen, H.L., 2003. Estimation of Monthly River Discharge from Danish Catchments Background and Objective of the Study. Nord. Hydrol. 34, 295–320.

Nash, E., Sutcliffe, V., 1970. River flow forecasting through conceptual models part i- a discussion of principles. J. Hydrol. 10, 282–290.

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

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

Parajka, J., Merz, R., Blöschl, G., 2005. A comparison of regionalization methods for catchment model parameters. Hydrol. Earth Syst. Sci. Discuss. 2, 509–542. https://doi.org/10.5194/hessd-2-509-2005.

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.

Peel, M.C., Finlayson, B.L., McMahon, T.A., 2007. Updated world map of the Köppen-Geiger climate classification. Hydrol. Earth Syst. Sci. 11, 1633–1644.

Perrin, C., Michel, C., Andreassian, V., 2001. Does a large number of parameters enhance model performance? Comparative assessment of common catchment model structures on 429 catchments. J. Hydrol. 242, 275–301.

Perrin, C., Michel, C., Andréassian, V., 2003. Improvement of a parsimonious model for streamflow simulation. J. Hydrol. 279, 275–289. https://doi.org/10.1016/S0022-1694(03)00225-7.

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.

Poissant, D., Arsenault, R., Brissette, F., 2017. Impact of parameter set dimensionality and calibration procedures on streamflow prediction at ungauged catchments. J. Hydrol. Reg. Stud. 12, 220–237. https://doi.org/10.1016/j.ejrh.2017.05.005.

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

Pushpalatha, R., Perrin, C., Le Moine, N., Andréassian, V., 2012. A review of efficiency criteria suitable for evaluating low-flow simulations. J. Hydrol. 420–421, 171–182.

Razavi, T., Coulibaly, P., 2013. Streamflow prediction in ungauged basins: review of regionalization methods. J. Hydrol. Eng. 18, 958–975. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000690.

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, 1–15. https://doi.org/10.1029/2008WR007248.

Rojas-Serna, C., Lebecherel, L., Perrin, C., Andréassian, V., Oudin, L., 2016. How should a rainfall-runoff model be parameterized in an almost ungauged catchment? A methodology tested on 609 catchments. Water Resour. Res. 1–24. https://doi.org/10.1002/2016WR018704.Received.

Salinas, J.L., Laaha, G., Rogger, M., Parajka, J., Viglione, A., Sivapalan, M., Blöschl, G., 2013. Comparative assessment of predictions in ungauged basins-Part 2: Flood and low flow studies. Hydrol. Earth Syst. Sci. 17, 2637–2652. https://doi.org/10.5194/hess-17-2637-2013.

Samuel, J., Coulibaly, P., Metcalfe, R.A., 2011. Estimation of continuous streamflow in ontario ungauged basins: comparison of regionalization methods. J. Hydrol. Eng. 16, 447–459. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000338.

Seibert, J., Beven, K.J., 2009. Gauging the ungauged basin: how many discharge measurements are needed? Hydrol. Earth Syst. Sci. 13, 883–892. https://doi.org/10.5194/hessd-6-2275-2009.

Seibert, J., Vis, M.J.P., 2012. Teaching hydrological modeling with a user-friendly catchment-runoff-model software package. Hydrol. Earth Syst. Sci. 16, 3315–3325. https://doi.org/10.5194/hess-16-3315-2012.

Shuttleworth, W.J., 1993. Evaporation. In: Maidment, D.R. (Ed.), Handbook of Hydrology. McGraw-Hill, New York, pp. 4.1–4.53.

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

Tveito, O.E., Bjørdal, I., Skjelvåg, A.O., Aune, B., 2005. A GIS-based agro-ecological decision system based on gridded climatology. Meteorol. Appl. 12 (1).

Valéry A. 2010. Modélisation précipitations–débit sous influence nivale.Élaboration d'un module neige et évaluation sur 380 bassins versants. Agro Paris Tech: Paris, France.

Vandewiele, G.L., Xu, C.Y., Huybrechts, W., 1991. Regionalization of physically-based water balance models in Belgium. Application to ungauged catchments. Water Resour. Manag. 5, 199–208. https://doi.org/10.1007/BF00421989.

Vaze, J., Post, D.A., Chiew, F.H.S., Perraud, J., Viney, N.R., Teng, J., 2010. Climate non-stationarity – Validity of calibrated rainfall – runoff models for use in climate change studies. J. Hydrol. 394, 447–457. https://doi.org/10.1016/j.jhydrol.2010.09.018.

Viglione, A., Parajka, J., Rogger, M., Salinas, J.L., Laaha, G., Sivapalan, M., Blöschl, G., 2013. Comparative assessment of predictions in ungauged basins - Part 3: Runoff signatures in Austria. Hydrol. Earth Syst. Sci. 17, 2263–2279. https://doi.org/10.5194/hess-17-2263-2013.

Viney, N.R., Perraud, J., Vaze, J., Chiew, F.H.S., Post, D.A., Yang, A., 2009a. The usefulness of bias constraints in model calibration for regionalization to ungauged catchments. 18th World IMACS/MODSIM Congr. Cairns, Aust. 3421– 3427.

Viney, N.R., Vaze, J., Chiew, F.H.S., Perraud, J.M., Post, D.A., Teng, J., 2009b. 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, pp. 3428–3434.

Vormoor, K., Lawrence, D., Schlichting, L., Wilson, D., Kwok, W., 2016. Evidence for changes in the magnitude and frequency of observed rainfall vs. snowmelt driven floods in Norway. J. Hydrol. 538, 33–48. https://doi.org/10.1016/j.jhydrol.2016.03.066.

Widén-Nilsson, E., Halldin, S., Xu, C., y.,, 2007. Global water-balance modelling with WASMOD-M: Parameter estimation and regionalization. J. Hydrol. 340, 105–118. https://doi.org/10.1016/j.jhydrol.2007.04.002.

Xu, C., 1999a. Operational testing of a water balance model for predicting climate change impacts. Agric. Forest Meteorol. 23, 95–304.

Xu, C., Halldin, S., 1997. The Effect of Climate Change on River Flow and Snow Cover in the NOPEX Area Simulated by a Simple Water Balance Model. Nordic Hydrol. 28 (4/5), 273–282.

Xu, C., Singh, V.P., 2002. Cross Comparison of Empirical Equations for Calculating Potential Evapotranspiration with Data from Switzerland. Water Resour. Manag. 16, 197–219.

Xu, C.Y., 2003. Testing the transferability of regression equations derived from small subcatchments to a large area in central Sweden. Hydrol. Earth Syst. Sci. 7, 317–324. https://doi.org/10.5194/hess-7-317-2003.

Xu, C.Y., 1999b. Estimation of parameters of a conceptual water balance model for ungauged catchments. Water Resour. Manag. 13, 353–368. https://doi.org/10.1023/A:1008191517801.

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

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

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

Zhang, Y., Chiew, F., 2009. Evaluation of Regionalization Methods for Predicting Runoff in Ungauged Catchments in Southeast Australia. 18th. World IMACS/MODSIM Congr. Cairns Aust, 3442–3448.

Zhang, Y., Vaze, J., Chiew, F.H.S., Teng, J., Li, M., 2014. Predicting hydrological signatures in ungauged catchments using spatial interpolation, index model, and rainfall – runoff modelling. J. Hydrol. 517, 936–948. https://doi.org/10.1016/j.jhydrol.2014.06.032.

Zhang, Y., Zheng, H., Chiew, F.H.S., Arancibia, J.P., Zhou, X., 2016. Evaluating Regional and Global Hydrological Models against Streamflow and Evapotranspiration Measurements. J. Hydrometeor. 17, 995–1010. https://doi.org/10.1175/JHM-D-15-0107.1.

Zhang, Z.X., Chen, X., Xu, C.-Y., Yuan, L.F., Yong, B., Yan, S.F., 2011. Evaluating the non-stationary relationship between Precipitation and Streamflow in Nine Major Basins of China during the past 50 years. J. Hydrol. 409, 81–93.

Zhao, R.-J., 1992. Xinanjiang model applied in China. J. Hydrol. 135 (2), 371–381.

Zhao R-J, Zuang Y, Fang L, Liu X, Zhang Q (1980). The Xinanjiang Model. In: Hydrological Forecasting. IAHS Press, Wallingford, pp. 351–356.