J. Hydrol. Hydromech., 66, 2018, 4, 381–392
DOI: 10.2478/johh-2018-0031

Factors controlling alterations in the performance of a runoff model in changing climate conditions

Patrik Sleziak[1*], Ján Szolgay[1], Kamila Hlavčová[1], Doris Duethmann[2], Juraj Parajka[2], Michal Danko[3]

[1] Department of Land and Water Resources Management, Slovak University of Technology, Radlinského 11, 810 05, Bratislava, Slovakia.
[2] Institute of Hydraulic Engineering and Water Resources Management, Vienna University of Technology, Karlsplatz 13/222, A-1040, Vienna, Austria.
[3] Institute of Hydrology, Slovak Academy of Sciences, Ondrašovská 17, 031 05 Liptovský Mikuláš, Slovakia.
[*] Corresponding author. E-mail: patrik.sleziak@stuba.sk

Abstract: In many Austrian catchments in recent decades an increase in the mean annual air temperature and precipitation has been observed, but only a small change in the mean annual runoff. The main objective of this paper is (1) to analyze alterations in the performance of a conceptual hydrological model when applied in changing climate conditions and (2) to assess the factors and model parameters that control these changes. A conceptual rainfall-runoff model (the TUW model) was calibrated and validated in 213 Austrian basins from 1981–2010. The changes in the runoff model's efficiency have been compared with changes in the mean annual precipitation and air temperature and stratified for basins with dominant snowmelt and soil moisture processes. The results indicate that while the model's efficiency in the calibration period has not changed over the decades, the values of the model's parameters and hence the model's performance (i.e., the volume error and the runoff model's efficiency) in the validation period have changed. The changes in the model's performance are greater in basins with a dominant soil moisture regime. For these basins, the average volume error which was not used in calibration has increased from 0% (in the calibration periods 1981–1990 or 2001–2010) to 9% (validation period 2001–2010) or –8% (validation period 1981–1990), respectively. In the snow-dominated basins, the model tends to slightly underestimate runoff volumes during its calibration (average volume error = –4%), but the changes in the validation periods are very small (i.e., the changes in the volume error are typically less than 1–2%). The model calibrated in a colder decade (e.g., 1981–1990) tends to overestimate the runoff in a warmer and wetter decade (e.g., 2001–2010), particularly in flatland basins. The opposite case (i.e., the use of parameters calibrated in a warmer decade for a colder, drier decade) indicates a tendency to underestimate runoff. A multidimensional analysis by regression trees showed that the change in the simulated runoff volume is clearly related to the change in precipitation, but the relationship is not linear in flatland basins. The main controlling factor of changes in simulated runoff volumes is the magnitude of the change in precipitation for both groups of basins. For basins with a dominant snowmelt runoff regime, the controlling factors are also the wetness of the basins and the mean annual precipitation. For basins with a soil moisture regime, landcover (forest) plays an important role.

Keywords: Climate change; Efficiency of runoff model; TUW model; Regression trees; Austria.

INTRODUCTION

Conceptual rainfall-runoff (r-r) models are useful tools for a wide range of tasks such as flood forecasting (Nester et al., 2016), reservoir and water management (Farkas et al., 2016), climate studies (Merz et. al., 2011), etc. These models are simplifications of the complex processes of runoff generation in a catchment. Components of these models often have to be described by empirical functions based on observations. The models therefore usually contain a number of parameters that do not directly represent measurable entities and thus must be estimated through the model's calibration in order to adjust the behavior of the model to mimic the behavior of an actual system (Valent and Szolgay, 2012). In their practical use such models may be operated under conditions different from those used for the model's calibration and validation, e.g., when we need to simulate the streamflow caused by extreme meteorological conditions or under a changing climate. It is therefore of interest to evaluate a model's performance in situations when the model has to simulate runoff under conditions dissimilar to those observed during the model's calibration (Seibert, 2003) in order to provide trustworthy runoff simulations when running models under conditions that may be significantly different from those used for their calibration. There are still many unknowns concerning the actual extrapolation capacity of particular hydrological models, which in general may depend on the quality and availability of the inputs, the calibration period of the climate model, the degree of the identifiability (sensitivity) of the model parameters, and other factors (Coron et al., 2011). There is also a growing interest in rainfall-runoff modelling over larger spatial domains in a multi-basin manner in order to explore spatial patterns by methods of comparative hydrology (Pechlivanidis and Arheimer, 2015).

A model's performance, for which the term "model efficiency" will also be used throughout this study, can be quantified by many characteristics, such as the runoff volume error, Nash-Sutcliffe efficiency, peak flow errors, the error in the timing of the peaks, etc. (Beven, 2005). The study of model efficiency is important for various reasons; for example, it is important to know how reliable the streamflow and flood forecasts will be, and it is essential to know what the limits of the model efficiency are.

In this paper, we will focus on two aspects of the robustness of the TUW r-r model: the changing climatic conditions (compared to those of the model's calibration) and diverse physiographic/climatic zones (the multi-basin behavior of the model) on the territory of Austria. Both aspects are intended to shed light on the applicability of the model: the first aspect mainly focuses on climate studies, since the model's performance should remain good under climatically-different simulation periods at the application stage, when the models are used in a climate change context. The second should complement the knowledge gained from modelling a single catchment, and it also substitutes here for an investigation of the uncertainties of the at-site parameters by a spatial comparative analysis in a large number of catchments. The final aim is to detect links between the model's performance and physiographical characteristics in order to understand the inadequacies and strengths of the model's performance for its future use in Austria (and in similar settings).

It is well documented that the efficiency of many models for a given regional application domain depends on the quality of the data (e.g., Andréassian et al., 2001; Coron et al., 2014; Oudin et al., 2006; Perrin et al., 2008), the length of the calibration period (Brath et al., 2004; Coron et al., 2014; Perrin et al., 2007), and the model's structure (Bai et al., 2015; Das et al., 2008; Fenicia et al., 2011; Merz et al., 2011; Perrin et al., 2001, 2003; Valent and Szolgay, 2012; van Esse et al., 2013; Vaze et al., 2010). In this context, quite a few authors have investigated the relationships between the model's efficiency and the climatic and catchment characteristics (e.g., Brigode et al., 2013; Coron et al., 2012; Fowler et al., 2016; Magand et al., 2015; Merz et al., 2009, 2011; Nester et al., 2011; Osuch et al., 2015; Oudin et al., 2006; Seifert et al., 2012; Seiler et al., 2012; Sleziak et al., 2016a, b; Vaze et al., 2010). Nester et al. (2011) calibrated and validated a semi-distributed hydrological model for a set of 57 catchments in Austria and Germany and analyzed how the model efficiency related to various catchment and climatic characteristics. They found a relation between model efficiency (as evaluated by the Nash-Sutcliffe efficiency NSE) and climate characteristics (such as the mean annual precipitation and runoff) and catchment characteristics (the catchment's size). According to their study, the catchment's size was the most significant control on the model efficiency. This is consistent with previous studies (e.g., Das et al., 2008; Merz et al., 2009; van Esse et al., 2013), which showed that model efficiency is mainly controlled by a catchment's size. Along a similar line Bai et al. (2015) evaluated the efficiency (in terms of the NSE and the water balance error) of 12 hydrological models in relation to different catchment and climatic characteristics. According to their study, the model efficiency was mainly influenced by the aridity index, the variability of the runoff, and the catchment's area. Oudin et al. (2006) used two lumped hydrological models (GEJ4 and TOPMODEL) for a sample of 12 US catchments and evaluated their efficiency (in terms of the NSE and the water balance error). They found that the model efficiency was controlled by climatic indicators (mainly precipitation). Similar results have been reported by Vaze et al. (2010), who calibrated four r-r models (SIMHYD, Sacramento, MARG, IHACRES) for 61 catchments in Australia and evaluated their efficiency. They found that the annual precipitation was the main control on the model efficiency.

Investigations of the efficiency of models during periods with contrasting climates has been a frequent topic of recent scientific literature as demonstrated in a review of common problems linked with using r-r models by Coron et al. (2011). Vaze et al. (2010) found that the model efficiency decreases and the bias increases with differences in the annual precipitation between the calibration and validation periods. This is in agreement with Coron et al. (2012), who assessed the efficiency of three lumped r-r models (GRJ4, MORDOR6, and SIMHYD) in relation to various climatic characteristics. The results from their study showed that the model efficiency was mainly affected by changes in the mean rainfall between the calibration and validation periods. Numerous studies have found that the loss of model efficiency can be related to a difference in precipitation (e.g., Coron et al., 2012; Fowler et al., 2016; Oudin et al., 2006; Vaze et al., 2010). There is also evidence that prolonged dry conditions can lead to a degradation in model efficiency over time (Saft et al., 2015; Saft et al., 2016).

Several authors have observed a decreasing trend in model efficiency when the model parameters were transferred over time (Fowler et al., 2016; Merz et al., 2011; Osuch et al., 2015). For example, Merz et al. (2011) found errors in simulations after transferring parameters between climatically different periods. They also revealed significant correlations between model parameters [snow correction factor (SCF), degree-day factor (DDF), the BETA parameter of runoff generation and field capacity (FC)], and climatic characteristics (mean annual precipitation, mean annual air temperature).

One of the possible approaches for evaluating which climatic and catchment characteristics control a model's performance is to use the technique of regression trees. In recent years, this technique has gained great popularity as has been demonstrated in various modeling studies, e.g., Iorgulescu and Beven (2004), Stauer et al. (2010), Seibert et al. (2016), Kuentz et al. (2016), Poncelet et al. (2017). Poncelet et al. (2017) investigated catchment controls on daily runoff simulations in France, Germany, and Austria. They examined how catchment features (i.e., 29 climate and catchment characteristics) and model efficiency criteria (i.e., the Nash-Sutcliffe efficiency, the Kling-Gupta efficiency on inverse streamflow and the mean and deviation biases) are linked. The results from this study showed that the catchment features that most affect a model's performance were the flashiness of precipitation and streamflow, the seasonality of evaporation, and catchment's aridity. While several of the previous studies analyzed the relationship between the hydrological model efficiency and various climatic and catchment characteristics, in this study we intend to focus on a better quantification of the factors that control change in the hydrological model efficiency over time in the context of a case study by using the TUW r-r model over the whole territory of Austria.

The aim of the paper is to evaluate the temporal changes of the efficiency and the factors controlling these changes. In particular, two aspects of the model efficiency are studied, i.e., the effect of the temporal change of at-site climatic conditions as expressed by a) the mean catchment precipitation b) and the air temperature (compared to those of the model's calibration) in two large groups of catchments representing diverse physiographic/climatic zones (the multi-catchment behavior of the model). The two disjunctive catchment groups are delineated by a study of the sensitive parameters of a model and represent catchments with predominantly snow and rain-fed runoff, respectively. Specifically, we address two scientific questions: (1) What factors (i.e., climatic and catchment characteristics) control the temporal changes in the hydrological model efficiency in both groups? (2) To what extent is it possible to quantify and evaluate these factors? We have used a multi-dimensional analysis by regression trees to evaluate the factors controlling changes in the hydrological model efficiency. This methodology has been tested for 213 catchments, which provide a representative (e.g., Viglione et al., 2013) sample of the different climatic and physiographic conditions in Austria.

The paper is organized as follows: the hydrological model, the calibration strategy, and the regression tree technique are described in Section 2; the study area and catchment set are presented in Section 3; the results are presented and discussed in Section 4; and an overall discussion and conclusions are given in Section 5.

METHODS

The TUW model

Description of the TUW model

A lumped conceptual r-r model at the basin scale (the TUW model, Viglione and Parajka, 2014) was used for the modeling. This model has also previously been used in various modeling studies across Austria (e.g., Parajka and Blöschl, 2008; Parajka et al., 2007; Sleziak et al., 2016a, b; Viglione et al., 2013) and Europe (Ceola et al., 2015). The TUW model has a structure similar to the Hydrologiska Byråns Vattenbalansavdelning (HBV) model (Bergström, 1995). The model works on a daily time step and requires the following catchment inputs: daily precipitation totals, mean daily air temperature, and daily potential evapotranspiration.

The model has 15 calibration parameters (Table 1), and its structure can be divided into three routines: a snow, soil moisture, and runoff routine (Merz and Blöschl, 2004). The role of the snow routine is the accumulation and melting of snow by a degree-day concept, using degree-day factor (mm/°C day) (DDF) and melt air temperature parameter Tm (°C). The catch deficit of precipitation gauges during snowfall is corrected by a snow correction factor SCF (–). A threshold temperature interval Tr – Ts (°C) is used to distinguish between rainfall, snowfall and mix of rain and snow (e.g., Parajka et al., 2005).

The soil moisture routine represents the runoff generation in a basin. Its role is to simulate the processes taking place in the basin's soil profile. This routine includes parameters such as the BETA (–) parameter of runoff generation, the maximum soil moisture storage (FC) (mm), and the limit for potential evapotranspiration Lprat (–).

The runoff routine is used to transform the outflow from upper and lower reservoirs. This routine contains five parameters: parameters related to surface and subsurface runoff (k0, k1, and k2), the threshold storage state (Lsuz) (mm), the constant percolation rate (Cperc) (mm/day), the maximum base and low flows (Bmax) (day), and the Croute transformation parameter (day^2/mm). More details about the model and its structure are given, e.g., in Parajka et al. (2007).

Table 1. The TUW model parameters including lower and upper bounds. The parameter ranges were taken from the literature (e.g. Merz et al., 2011).

Parameter	Abbreviation, unit	Range
Snow correction factor	SCF (–)	0.9–1.5
Degree-day factor	DDF (mm/°C day)	0–5
Rain threshold temperature	Tr (°C)	1–3
Snow threshold temperature	Ts (°C)	–3–1
Melt temperature	Tm (°C)	–2–2
Limit for potential evapotranspiration	Lprat (day)	0–1
Maximum soil moisture storage	FC (mm)	0–600
Nonlinearity parameter	BETA (–)	0–20
Very fast storage coefficient	k0 (days)	0–2
Fast storage coefficient	k1 (days)	2–30
Slow storage coefficient	k2 (days)	30–250
Upper storage coefficient	LSuz (mm)	1–100
Percolation rate	Cperc (mm/day)	0–8
Maximum base parameter	Bmax (days)	0–30
Free scaling parameter	Croute (day^2/mm)	0–50

Calibration strategy

The TUW model was calibrated for three 10-year periods between 1981–2010. The model's parameters are estimated by automatic calibration using the Deoptim differential evolution algorithm (Ardia et al., 2015). This algorithm has also successfully been used in previous modeling studies (e.g., Sleziak et al., 2016a, b). Deoptim belongs to the class of genetics algorithms which use biology-inspired process such as crossover, mutation, and selection on a population. The principle of this algorithm is based on repeated evaluation of the objective function in order to move the population toward a global minimum. For more detailed information regarding the Deoptim algorithm see, e.g., Ardia et al. (2015).

The parameter ranges used in this study were taken from the literature (see, e.g., Merz et al. 2011; Viglione et al., 2013). A warm-up period of one year is used in the calibrations. The objective function combines the Nash-Sutcliffe efficiency (NSE) (Nash and Sutcliffe, 1970) and the logarithmic Nash-Sutcliffe efficiency (logNSE) (Merz et al., 2011). While the NSE puts a greater emphasis on high flows, the logNSE puts a great emphasis on low flows. Mathematically, these criteria can be expressed as follows:

$$NSE = 1 - \frac{\sum_{i=1}^{n}\left(Q_{sim,i} - Q_{obs,i}\right)^2}{\sum_{i=1}^{n}\left(Q_{obs,i} - \overline{Q}_{obs}\right)^2} \tag{1}$$

$$\log NSE = 1 - \frac{\sum_{i=1}^{n}\left(\log(Q_{sim,i}) - \log(Q_{obs,i})\right)^2}{\sum_{i=1}^{n}\left(\log(Q_{obs,i}) - \log(\overline{Q}_{obs})\right)^2} \tag{2}$$

where $Q_{sim,i}$ and $Q_{obs,i}$ represent the simulated and observed mean daily flows on day i, and $\overline{Q}_{obs}$ is the average of the flows observed. The NSE and logNSE coefficients range between $-\infty$ and 1 (NSE = 1 indicates a perfect simulation, i.e., an absolute agreement between the observed and simulated flows). The combination of NSE and logNSE in the objective function (*OF*) is defined as:

$$OF = \frac{1 - NSE}{2} + \frac{1 - \log NSE}{2} \tag{3}$$

The given objective function (Eq. 3) is minimized by the differential evolution algorithm Deoptim (see above).

Performance assessment

In this study, the Differential Split-Sample Test (DSST, Klemeš, 1986) has been adopted to evaluate the ability of the TUW model to simulate runoff under conditions different from those used for the model's calibration. The DSST allows us to evaluate the model's efficiency (performance) in contrasting climatic periods (i.e., in cooler and warmer periods or wetter and drier periods). Some applications of DSST can also be found in, e.g., Seibert (2003), Wilby (2005), Chiew et al. (2009), Vaze et al. (2010), Merz et al. (2011), Brigode et al. (2013), Bai et al. (2015), Sleziak et al. (2016a, b).

The quality of the model simulations is quantified by the Nash-Sutcliffe efficiency (NSE, Eq. 1) and the volume error (VE) (e.g., Merz et al., 2011). The VE is not used in calibration, this metric is used only for the assessment of the model performance. The value VE = 0 means that the simulated mean runoff equals the observed one (i.e., no bias). The values VE < 0 and

VE > 0 denote underestimations and overestimations of the flows.

Multi-dimensional analysis

In this study we applied a multi-dimensional approach using regression trees to identify factors that control changes in volume error between the calibration and validation periods. Here we chose VE values as representative for the comparison because they can be directly comparable between periods and catchments (e.g., Coron et al., 2012). A regression tree is an analytical and visualization method which, via tree-building algorithms, provides information about interactions between various characteristics (i.e., between the variables and the model's features). Based on this technique we can determine the importance of the explanatory variables and also find some common patterns. The principle of this technique is based on the binary splitting of a dataset according to features (i.e., decision variables), which are automatically selected by the algorithm. By implementing the rpart package (Therneau et al., 2017), it is possible to apply one of these algorithms (CART, Breiman et al., 1984) in the R software environment (R Development Core Team, 2011). In general, the CART algorithm creates two main clusters which are further divided into subsets. In order to simplify the structure of the trees we set the minimum number of basins in each final cluster (leaf) to 25. These smaller trees are better suited for capturing general patterns and this way we avoid very complex trees that are difficult to interpret. For more detailed information about regression trees, see, e.g., Kuentz et al. (2016) and Poncelet et al. (2017).

We apply regression trees to identify factors that control changes in bias when a model's parameters are transferred from the calibration to the validation period. As explanatory factors we use (1) climatic characteristics, such as the mean annual precipitation (P), the mean annual air temperature (T), the relative change in mean annual precipitation (Pdif), and the absolute change in mean annual air temperature (Tdif) and, (2) catchment characteristics such as the area, elevation, slope, percentage of forest cover, and aridity (defined as the ratio of the mean annual potential evapotranspiration to the mean annual precipitation).

STUDY REGION AND DATA

Austria was selected as the study region. This region was selected as the test bed region due to (a) the variability of the climate (i.e., an increase in precipitation and air temperatures over recent decades has been observed, Merz et al. 2011), (b) diverse physiographic conditions (i.e., different catchment areas, elevations, geology, etc., Gaál, et al. 2012), (c) the availability of inputs (i.e., precipitation, air temperatures, potential evapotranspiration and streamflow), and the suitability of inputs (i.e., quality of the data) for modeling experiments (e.g., Viglione et al., 2013).

We used a representative sample of 213 catchments in Austria, which represent a large variety of climatic and physiographic conditions of Central Europe. The catchment areas range from 14 to 6200 km^2; mean elevation range from 295 to 2915 m a.s.l. Annual precipitation varies from 400 to 3000 mm/year and mean annual air temperature from –8 to 10°C.

We used daily hydrometeorological catchment data (i.e., the daily precipitation totals, mean daily air temperature, mean daily streamflow, and daily potential evapotranspiration) from the period 1981–2010. These data have also been extensively used in previous modeling studies, e.g., Viglione et al. (2013), Sleziak et al. (2016a, b). Before processing the data, quality flags, missing data, etc., were visually inspected. The precipitation data came from 1091 rainfall gauges. These measurements were used to interpolate the catchments' mean areal precipitation using the external drift kriging method (see Merz et al., 2011). The air temperature data came from 212 climatic stations. The catchments' mean air temperatures were calculated using the least squares trend prediction method (Pebesma, 2001). The mean daily runoff data from 213 gauged catchments were provided by the Austrian Hydrographic Service (HZB). These data were used to calibrate and validate the TUW model. The daily potential evapotranspiration was calculated by a modified Blaney-Criddle method (Parajka et al., 2005). More details about the data (e.g., the methods used to interpolate the data) can be found, e.g., in Merz et al. (2011) and Viglione et al. (2013).

Classification of the basins based on a sensitivity analysis

The selected Austrian basins (213) were classified into two groups using a sensitivity analysis of the TUW model's parameters. This analysis was carried out using a combination of the Latin Hypercube (LH) and one-factor-at-a-time sampling (van Griensven et al., 2006), which allowed us to identify patterns of parameter dominance. The sensitivity analysis is used to detect the most relevant model parameters in relation to the objective functions selected (see calibration strategy, above). The LH subdivides the range of each parameter into N segments, each with a probability of occurrence equal to 1/N. Random values for each parameter are generated such that each of the segments are sampled one time. A detailed description of this method is presented in van Griensven et al., 2006.

In our case, the sensitivity of the model parameters was tested in three specific periods (i.e., 1981–1990, 1991–2000, and 2001–2010). This is documented in Table 2, which gives information about the occurrence of the most sensitive model parameters in these periods. For example, we can see that the degree-day melt parameter (DDF) was 87 times the most sensitive in 1981–1990, 65 times in 1991–2000, and 75 times in 2001–2010 (the value of 0 in the table means that the given parameter did not appear to be the most sensitive in a particular period). With this analysis we identified four model parameters

Table 2. Frequencies of the most sensitive model parameters in three specified calibration periods (1981–1990, 1991–2000, 2001–2010).

Parameter	1981–1990	1991–2000	2001–2010
SCF (–)	0	0	0
DDF (mm/°C day)	87	65	75
Tr (°C)	0	0	0
Ts (°C)	0	0	0
Tm (°C)	0	0	0
Lprat (day)	0	0	0
FC (mm)	91	127	130
BETA (–)	9	2	0
k0 (days)	0	0	0
k1 (days)	0	0	0
k2 (days)	0	0	0
LSuz (mm)	0	0	0
Cperc (mm/day)	20	19	8
Bmax (days)	0	0	0
Croute (day^2/mm)	0	0	0

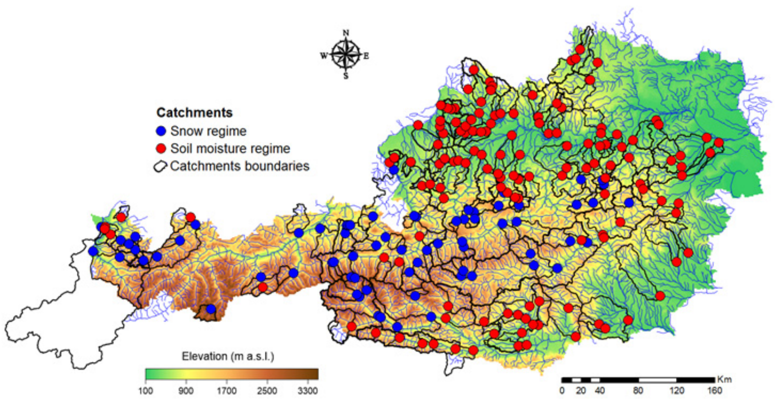

Fig. 1. Map of Austria with the selected catchments (213). The blue circles indicate catchments with a dominant snow regime (71 catchments), and the red circles are catchments with a dominant soil moisture regime (142 catchments). Both groups of catchments were delineated by a model parameter sensitive study.

which have a close relationship with the objective function selected. These results are also presented in Fig. 1 where we can identify two distinct regions showing different sensitivities to the model parameters. In alpine (mountainous) regions, the most sensitive parameters are related to the accumulation and melting of snow (degree-day melt parameter (DDF)). On the contrary, the most sensitive in the flatlands are the parameters related to the soil (maximum soil field capacity (FC), the BETA parameter related to runoff generation, and the percolation rate (Cperc)). These groups were also compared in terms of the hydrological regime. Using this comparison, we classified the 213 Austrian catchments into two main groups of snow (71 catchments, where the DDF was the most sensitive parameter) and soil moisture regimes (142 catchments, where FC, BETA or Cperc was the most sensitive parameter). Table 3 summarizes the main characteristics of these two groups of basins.

Table 3. Basic statistical values (i.e., minimum, median, maximum and mean) computed for the selected characteristics of two group of basins, i.e., basins with a) dominant snow and b) soil moisture regimes.

Area (km^2)	**Min**	**Median**	**Max**	**Mean**
Basins with a snow regime	14.2	150.8	6120	594.8
Basins with a soil moisture regime	13.7	178.2	6214	319.0
Mean elevation (m a.s.l.)	**Min**	**Median**	**Max**	**Mean**
Basins with a snow regime	984	1551	2915	1636
Basins with a soil moisture regime	295	702	1924	795
Forest cover (%)	**Min**	**Median**	**Max**	**Mean**
Basins with a snow regime	2.1	40.3	86.8	43.8
Basins with a soil moisture regime	9	56	98	55.6
Slope (%)	**Min**	**Median**	**Max**	**Mean**
Basins with a snow regime	19.9	38.9	54.2	38.8
Basins with a soil moisture regime	2.6	16.1	47.8	19.3

For a better understanding of the variability between the two different groups of catchments (i.e., catchments with snow and catchments with a soil moisture regime) and also for a better interpretation of the temporal changes in the model parameters, we plotted the changes in the climatic characteristics over three decades for both groups separately (Fig. 2). Figure 2 shows that catchments with a snow-dominant regime are characterized by higher precipitation and lower air temperatures. While the median values of the mean annual precipitation (P) increased from 1032 to 1095 mm/year in catchments with a dominant soil moisture regime, catchments with a dominant snow regime have fairly stable median values (between 1460 and 1455 mm/year). Also, the medians of the mean annual air temperatures (T) show an increasing trend over three decades. The median values of the mean annual air temperatures increased on average by 0.7°C and 0.4°C in catchments with snow and soil moisture regimes, respectively. Interestingly, the medians of the mean annual runoff (Q) practically did not change over three decades.

Based on this analysis, it would be interesting to examine how these changes in climatic characteristics can be linked to changes in the model parameters.

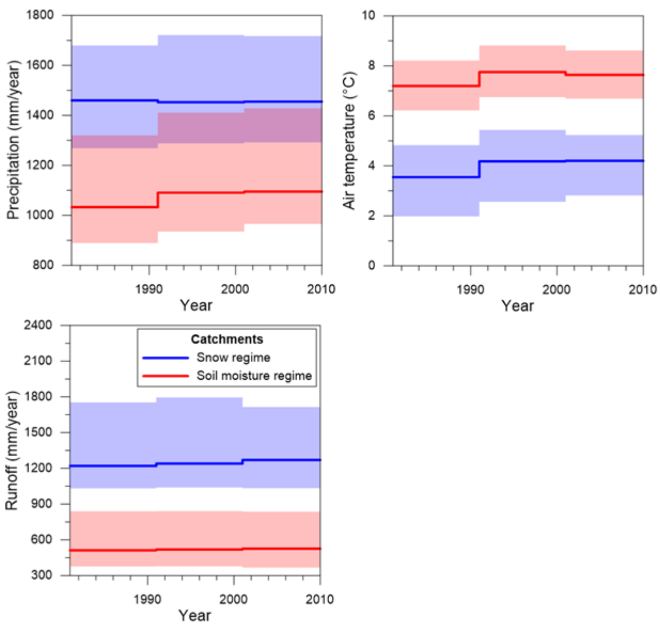

Fig. 2. Changes in the hydroclimatic characteristics (mean annual precipitation, air temperature and runoff) over three decades. The blue line indicates medians for catchments with a snow regime (71 catchments), and the red line shows medians for the catchments with a soil moisture regime (142 catchments). The shaded area represents 75% and 25% percentiles between the catchments.

RESULTS AND DISCUSSION

Assessment of the model's efficiency in different climatic periods

In this section we assess the model's efficiency over different climatic periods (i.e., colder/warmer or drier/wetter periods) for the two groups of catchments. In this study, the period 1981–1990 is considered as colder/drier and the period 2001–2010 as wetter/warmer (see section above).

In the snow-dominated catchments, the mean NSE values during the calibration period were between 0.73 and 0.74 and varied only slightly over the decades (Fig. 3), which means that the model could be calibrated equally well in each period. The average volume errors of –0.02 to –0.07 indicate a tendency to underestimate the simulated runoff volume in the calibration as well as the validation periods.

In the catchments with a dominant soil moisture regime, the median NSE values during the calibration period were between 0.7 and 0.73 for the three calibration periods (Fig. 4), thereby showing a slightly lower model performance in comparison with the snow-dominated catchments. The decrease in the model performance from the calibration to the validation periods is greatest between the calibration from 2001–2010 and the validation from 1981–1990. The assessment of the VE efficiency indicates that the model's calibration is essentially unbiased (the median VE equals 0). Interestingly, the model tends to overestimate flows in warmer validation periods, but underestimates flows in colder and drier decades.

The comparison of the runoff model efficiencies in the two groups of catchments indicates larger NSE efficiencies in catchments with a dominant snow regime for both the calibration and validation periods.

In this context, Schaefli and Gupta (2007) pointed out the correct interpretation of the NSE values. The benchmark model when using NSE is the mean observed flow. In strongly seasonal regimes (like snow-dominated regimes with low flows during winter and high flows during the melting period), a model that can capture the general seasonal regime already achieves a good NSE and therefore often higher NSE values are obtained in snow dominated catchments.

Similar to our study, the Viviroli et al. (2009) focused on calibration of large number of catchments (i.e., 140 catchments in Switzerland) between different calibration periods. They calibrated the distributed r-r model PREVAH (Precipitation-Runoff-EVApotranspiration-HRU model) in hourly time step between 1984–2003. They proposed a robust calibration strategy that combines traditional calibration approach (multiple measures of goodness-of-fit) and a fuzzy approach (used for modeling high flows). They observed a decreasing trend in model performance after transferring parameters between contrasted calibration periods. For example, for 49 representative basins (i.e., with a long record of observations) the median of NSE decreased from 0.75 (calibration period) to 0.72 (validation period). For high flows slightly poorer values of NSE were obtained (i.e., 0.69 in calibration and 0.67 in validation).

Merz et al. (2011) found errors in simulations (overestimation of Q_{95} by about 12%, overestimation of Q_{50} by about 15%, and overestimation of Q_{05} by about 35%) when the model parameters were transferred over time.

Our findings also indicate that in snow-dominated catchments the model tends to systematically underestimate the volume of the flows. On the contrary, in catchments with a soil moisture regime, the model simulates flow volume closer to the observation in the calibration periods, but overestimates flows when the model parameters are transferred from colder/drier to wetter/warmer decades (i.e., from 1981 – 1990 to 2001-2010) . These results contrast with previous findings, e.g., Vaze et al. (2010) or Coron et al. (2012), who observed a tendency to overestimate mean runoff when the calibration period was wetter (i.e., a wet to dry parameter transfer). This can be related to the different regions (i.e., Austria vs Australia) and physiographic conditions studied.

Generally, for hydrological simulations under varying conditions, physically based distributed models are usually preferred over conceptual r-r models (like TUW), because of their process foundation they are valid for conditions outside the calibration period and also allow a better description of spatial heterogeneity (e.g., Finger et al., 2012). However, they require more input data, which is often not available, and have more parameters. Furthermore, the higher complexity of these models demands a longer computation time (Sun, et al., 2017). While conceptual models like HBV have minimal data requirements, require minimal computing time, they may not be well suited under changing conditions.

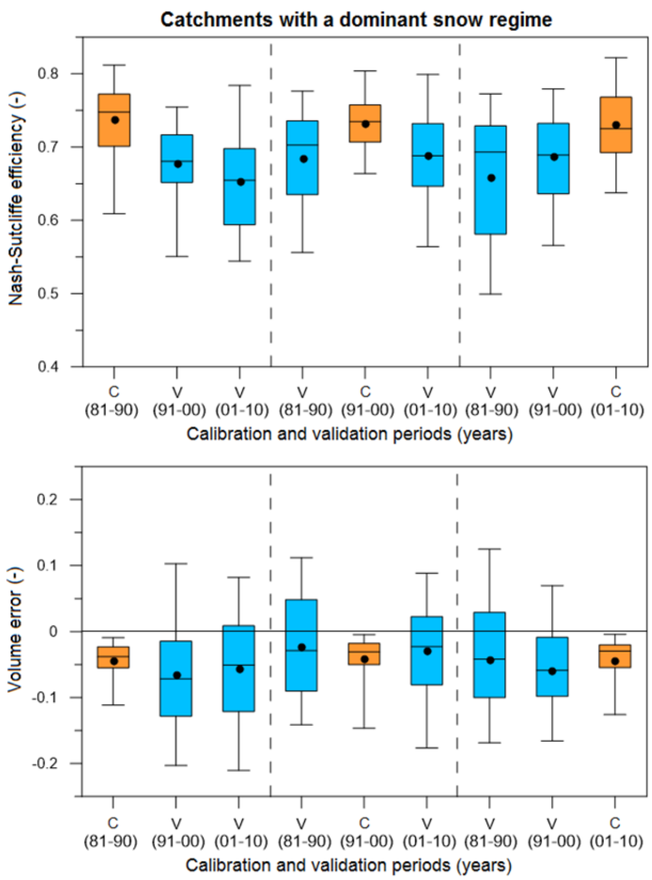

Fig. 3. Comparison of the variability of the Nash-Sutcliffe efficiency (NSE) and the volume error (VE) in the specified calibration and validation periods for catchments with a dominant snow regime. The horizontal line of box plots shows the median of the values, and the upper and lower whiskers show the 95 and 5 percentile values. The horizontal line inside the bottom graph shows a zero volume error. The orange box plots represent the calibration periods. The blue box plots represent the validation periods. The line inside the boxes shows the median of the NSE and VE. The black circles are the mean values of the NSE and VE. C are calibration periods and V are validation periods.

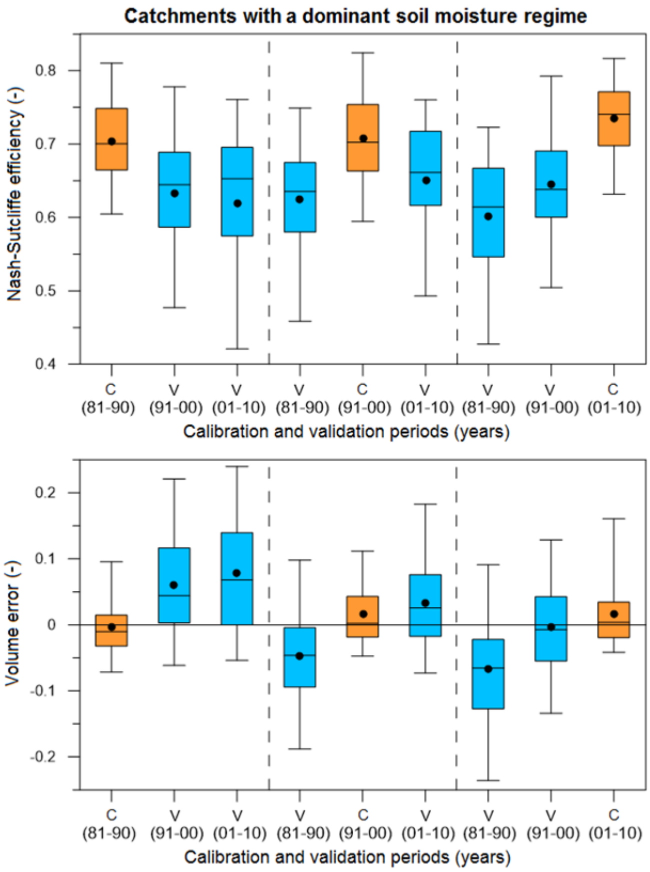

Fig. 4. Comparison of the variability of the Nash-Sutcliffe efficiency (NSE) and the volume error (VE) in the specified calibration and validation periods for catchments with a dominant soil moisture regime. The horizontal line of the box plots shows the median of the values, and the upper and lower whiskers show the 95 and 5 percentile values. The horizontal line inside the bottom graph shows a zero volume error. The orange box plots represent the calibration periods. The blue box plots represent the validation periods. The line inside the boxes shows the median of the NSE and VE. The black circles are the mean values of the NSE and VE. C are calibration periods, V are validation periods.

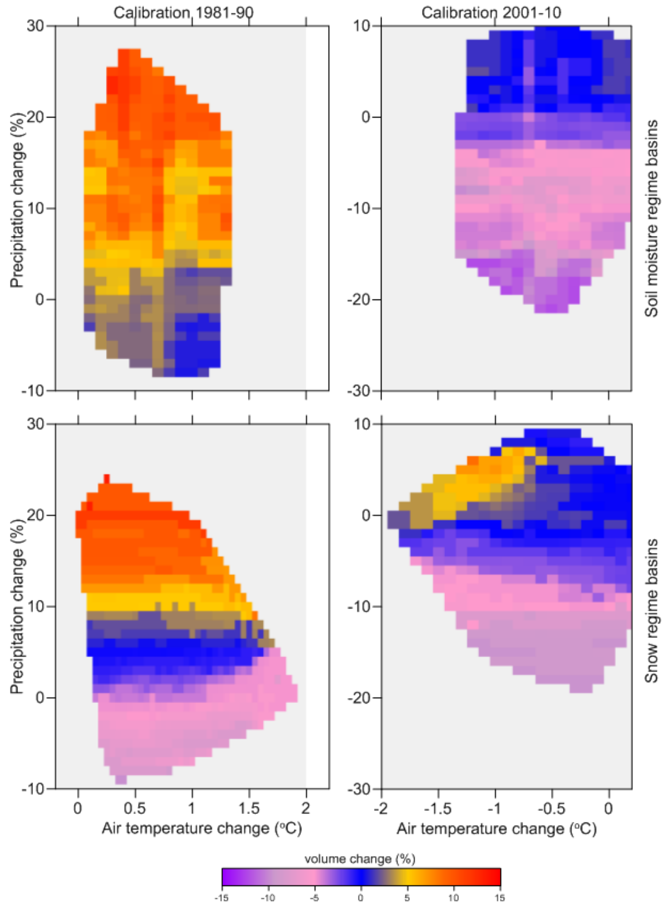

Fig. 5. Changes in the simulated runoff volume error when transferring model parameters calibrated in 1981–1990 (left panels) or 2001–2010 (right panels) to the remaining two decades. The changes in air temperature and precipitation indicate differences with respect to the calibration period. The change is estimated for basins with dominant snowmelt (bottom panels) and soil moisture (top panels) regimes.

The relationship between changes in the climatic characteristics and changes in the volume errors

Figure 5 shows changes in the simulated runoff volume errors between the calibration and the two validation periods as a function of changes in the mean annual precipitation and air temperature between those time periods. Thus, each catchment results in two data points. The data points were interpolated to obtain a smooth response surface. While the left panels in Figure 5 show the changes from a colder/drier calibration decade (i.e., 1981–1990) to warmer/wetter validation periods, the right panels show the changes from a warmer/wetter calibration period (i.e., 2001–2010) to colder/drier validation periods. The top and bottom panels show the changes in the snow and soil moisture-dominated catchments, respectively. The results indicate that transferring parameters to a colder/drier or warmer/wetter decade leads to runoff underestimation or overestimation, respectively. This trend is more pronounced in the catchments dominated by soil moisture processes. It can be seen that changes in simulated runoff volume are mainly related to changes in precipitation, but that this relationship tends not to be linear. An increase in mean annual precipitation of 10–20% leads to an increase in volume error of 5–15%. In contrast, when the change in precipitation is in a range of 0 to –20% and the air temperature decreases, the change in runoff volume errors is in a range of –5% to –10%. These results are in general agreement with the findings presented, e.g., by Oudin et al. (2006), Vaze et al. (2010), and Coron et al. (2012), who showed the significant effect of changing precipitation on a runoff model's efficiency. For example, Coron et al. (2012) compared the model efficiency of three commonly used lumped r-r models (GRJ4, MORDOR6, and SIMHYD) in relation to selected climatic characteristics (i.e., mean annual precipitation, air temperature). The results of their study indicated a 20% bias in total volumes when the mean rainfall differed by 10–20% between the calibration and validation periods. This study was performed in 216 catchments in southeast Australia. Along similar lines, Vaze et al. (2010) reported that simulations of runoff are acceptable when changes in precipitation are no more than 15% less or 20% greater than the precipitation during the calibration period.

Changes in the model parameters in different climatic periods

The changes in the selected model parameters (i.e., the degree-day melt parameter (DDF), the BETA parameter of runoff generation, the maximum soil moisture storage (FC), and the percolation rate (Cperc)) over three different decades are shown in Fig. 6. The selection of these parameters is based on the

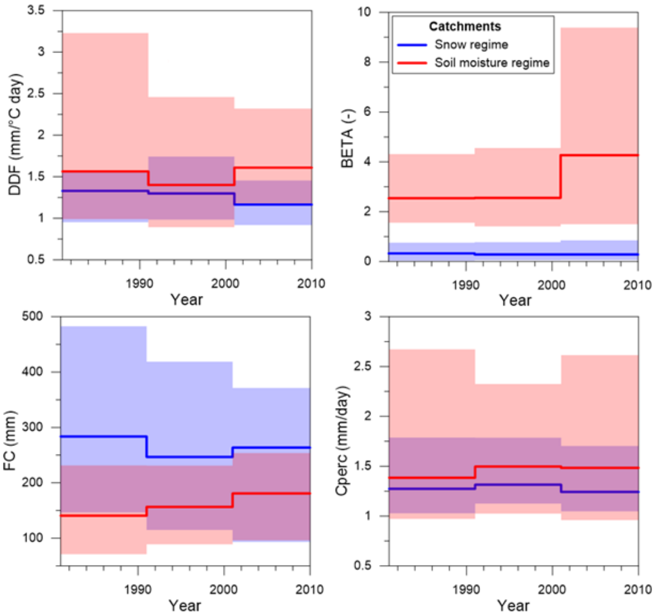

Fig. 6. Changes in the model parameters (the degree-day factor (DDF), the BETA parameter of runoff generation, the maximum soil moisture storage (FC) and the percolation rate (CPERC)) in three different climatic periods (1981–1990, 1991–2000, 2001–2010). The blue line indicates medians for the catchments with a snow regime (71 catchments), and the red line shows medians for catchments with a soil moisture regime (142 catchments). The shaded area represents 75% and 25% percentiles between the catchments.

results of the sensitivity analysis (see Section data, above). The changes in the parameters are described by selected percentiles (i.e., 25%, 50%, and 75%). The degree-day melt parameter (DDF) is one of the most sensitive parameters in catchments with a dominant snow regime. The results indicate that DDF tends to decrease in warmer/wetter decades. Such changes are likely to be associated with changes in snow melting in the spring, which tend to be greater in colder years (1981–1990) (see Merz et al., 2011). The DDF values in catchments with a soil moisture regime have a much larger scatter, but the median values are more stable over the decades. The BETA nonlinear runoff generation parameter and field capacity (FC) are the most sensitive parameters in catchments with a soil moisture regime. Both tended to increase over the decades analyzed (Fig. 6). This increase is likely associated with increasing evapotranspiration, mainly in flatland catchments. As was pointed out by Merz et al. (2011), an increasing trend in the BETA parameter is also connected with a more non-linear runoff generation in the last decade analyzed. The median values of FC in basins with a snow regime show a decreasing trend over decades; however, the scatter is large. Interestingly, the values of FC are higher in snow-dominated basins, which are largely in the mountainous region where one would expect shallow soils (generally, the small FC values imply shallow hydrologically active soil depths and vice versa, Merz et al., 2011). This is possibly due to (a) cross-correlation between the parameters BETA and FC, i.e., when FC is high, then BETA is small, (b) heterogeneous conditions (i.e., alpine vegetation, forests, bareland covers a substantial portion of the catchments areas, e.g., Merz and Blöschl, 2004). The findings also show that the median maximal available FC increased from 95 mm to approximately 150 mm within 20 years (Fig. 6). This can be related to increasing trend in air temperature (FC may vary in response to the variability of climatic conditions) between periods (see Fig. 2). The change in FC does not necessarily mean that the storage capacity of the soils has changed but can also be a sign of a compensation effect for the rather low potential evaporation (Nijzink et al., 2016). In this context, Wang-Erlandsson et al. (2016) pointed out that the root storage zone capacity (which determines the maximum soil moisture) is critical for correctly simulating surface runoff. Authors in their study present a method to estimate root zone storage capacity from satellite-based evaporation and observation-based precipitation data. The results showed that the method eliminated the need for poor resolution soil and rooting depth data and therefore can be useful for the modelling community.

The percolation rate (Cperc) is only sensitive in a few catchments, and the median values did not change much in both groups of catchments. Merz et al. (2011) found that changes in model parameters are related to changes in climate variables, such as increases in air temperature and potential evapotranspiration. Here, we show that the changes in model parameters are further related to whether the catchment belongs to a snow or soil processes-dominated group (i.e., different climatic conditions – catchments in mountainous part of regions vs. catchments in flatlands, see section Study region and data). In comparison to our study, Merz et al. (2011) analyzed changes in model parameters calibrated by a semidistributed version of the TUW model in different 5-year periods. Our results for DDF, the most sensitive snow model parameter, and the BETA nonlinear runoff generation parameter show similar trends in changes over time, but the tendency is different for the field capacity (FC) model parameter. While Merz et al. (2011) presented a clear increase in FC during 5-year calibration periods, our results indicate an increasing trend only in flatland catchments. The difference between the results is likely caused by the different spatial distribution of the model inputs and differences in the length of the calibration periods.

Factors controlling changes in runoff volume

In this section, we use regression trees to investigate which climatic and catchment characteristics controlled the differences in the simulated runoff volume error (VEdif) in two different decades.

The regression trees are identified separately for the two groups of catchments, and the results are presented in Figs. 7 and 8. Figure 7 shows a case of changing simulated flow volume (VEdif), when the model is calibrated in a colder/drier decade and applied (validated) in a warmer/wetter time decade. In both groups of catchments, the main controlling factor of the VEdif is the magnitude of the precipitation change (Pdif). In the snow-dominated catchments, an increase in the annual precipitation greater than 2.3% results in an average increase in the simulated runoff volume of 5%. In contrast, a decrease in the mean annual precipitation, particularly in catchments with a mean annual precipitation lower than 1834 mm and a size less than 124 km^2, results in an average 12% decrease in the simulated flow volume. In catchments with a dominant soil moisture regime (Fig. 7, right panels), a change in mean annual precipitation also affects the degree by which the simulated runoff volumes increase. In almost all the catchments the application of the model parameters from a colder/drier period to a warmer/wetter period leads to increased simulated runoff volumes. The largest increase was observed in catchments where the mean annual precipitation increased by at least 3.7% and the forest cover was greater than 45%.

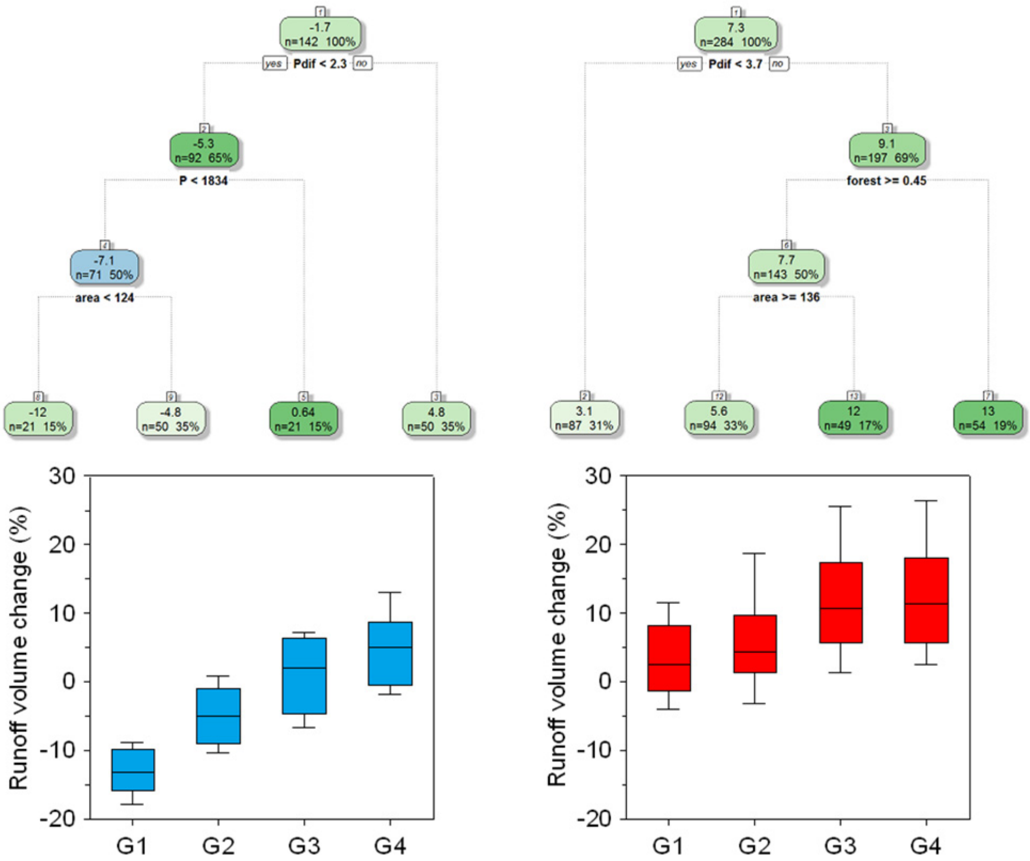

Fig. 7. Factors describing changes in runoff volume when simulating runoff in warmer and wetter decades (1991–2000 and 2001–2010) by using model parameters calibrated in a colder and drier period (1981–1990). The left side shows basins with a snow regime; the right side shows basins with a soil moisture regime. G1–G4 represent the final clusters (leaves) of the resulting trees.

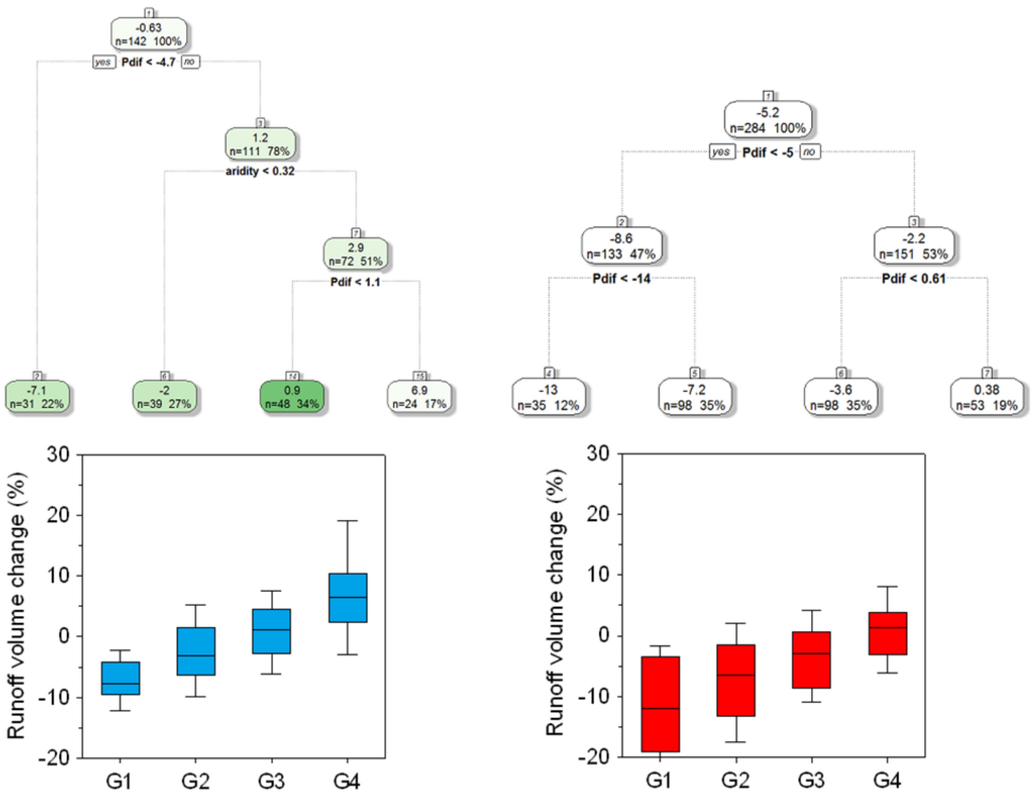

Fig. 8. Factors describing changes in runoff volume when simulating runoff in colder and drier decades (1981–1990 and 1991–2000) by using model parameters calibrated in a warmer and wetter period (2001–2010). The left side shows basins with a snow regime; the right side shows basins with soil moisture regime. G1–G4 represent the final clusters (leaves) of the resulting trees.

Fig. 8 shows the factors influencing the simulated runoff volume when applying model parameters calibrated in a warmer/wetter (2001–2010) decade and applied to colder/drier decade. The changes in the simulated runoff volume in different climate conditions are again mostly determined by changes in the mean annual precipitation (Pdif). The Pdif is the only differentiating factor in catchments with a dominant soil moisture regime (Fig. 8, right panels). Interestingly, in the snow-dominated catchments the second most influential factor is the aridity of the catchment. For drier catchments (aridity < 0.32) or catchments with a small Pdif, the difference in the changes in the simulated volume is the smallest (a range of –2% to 1%).

The results are in agreement with previous analyses (e.g., Coron et al., 2012; Oudin et al., 2006; Vaze et al., 2010), which showed that changes in model efficiency are mainly affected by changes in the mean annual precipitation. These results also correspond with the greatest correlations (not presented here) between the magnitude of changes in precipitation and changes in runoff volume. The size of the catchments and amount of aridity, which were determined to be important factors in previous studies (Nester et al., 2011; Poncelet et al., 2017), are only influential for smaller and snow-dominated catchments, respectively. In flatland catchments, (landcover) forest also has some impact on changes in the bias of simulated runoff volumes.

CONCLUSIONS

Hydrological models are simplified mathematical representations of complex rainfall-runoff processes. These models can be considered useful tools for the diagnosis of the impacts of climate change on water resources. The use of these models under conditions that may be significantly different to those used for their development still remains a challenging task.

In this study, we have evaluated the impact of changing climate conditions with differential split-sample testing. We found that changes in simulated runoff volume are clearly related to changes in precipitation, but the relationship is not always linear, particularly in flatland catchments. Our results indicate that a trustworthy simulation can be achieved if a change in mean precipitation is no more than ±10%.

The analysis of temporal changes in model parameters over three different decades showed that the parameters controlling snow processes (i.e., snow correction factor SCF and degree-day factor DDF), and soil moisture processes (i.e., parameter of runoff generation BETA and maximum soil moisture storage FC) have changed over time.

The evaluation of factors which control changes in simulated runoff volumes showed that the most influential factor is a change in mean annual precipitation. Additionally, the aridity or wetness of the catchments had some influence on catchments with a dominant snowmelt runoff regime.

From the results of our study we conclude that it is indeed important to re-calibrate conceptual r-r models (e.g., the HBV type and its derivatives) if the climatic conditions change. For practical applications of hydrological models, it would be suitable to consider various calibration periods and change the model parameters, depending on the hydroclimatic regime. The findings also revealed a need for regionalization of the methodology and a need to verify it in different climatic and physiographic conditions. More analysis needs to be done in the future, for example to apply similar testing approaches to different regions, to use HVB vs. physically based models, etc.

Acknowledgements. This work was supported by the Slovak Research and Development Agency under Contract Nos. APVV-15-0497 and APVV-15-0425; and VEGA 1/0710/15. We would like to thank the Austria Science Foundation (as part of the Vienna Doctoral Programme on Water Resource Systems (DK-plus W1219-N22) and the European Research Council under the ERC Advanced Grant "Flood Change", Project No. 291152. We gratefully acknowledge support from the German Research Foundation (DFG) through a research scholarship to D.D. At the same time, we would also like to acknowledge the financial support from the National Scholarship Programme of the Slovak Republic. The contribution is also the result of the project implementation ITMS 26240120004 Centre of excellence for integrated flood protection of land supported by the Research & Development Operational Program funded by the ERDF. The authors are grateful to Peter Molnar and also two anonymous reviewers whose valuable comments and constructive suggestions helped improve and clarify the paper.

REFERENCES

Andréassian, V., Perrin, C., Michel, C., Usart-Sanchez, I., Lavarbe, J., 2001. Impact of imperfect knowledge on the efficiency and the parameters of watershed models. Journal of Hydrology, 205, 1–4, 206–223. http://dx.doi.org/10.1016/S0022-1694(01)00437-1.

Ardia, D., Mullen, K.M., Peterson, B.G., Ulrich, J., 2015. DEoptim: Diferential evolution in R. Version 2.2-3.

Bai, P., Liu, X., Liang, K., Liu, C., 2015. Comparison of performance of twelve monthly water balance models in different climatic catchments of China. Journal of Hydrology, 529, 1030–1040. DOI: 10.1016/j.jhydrol.2015.09.015.

Bergström, S., 1995. The HBV model. In: Sing, V.P. (Ed.): Computers Models of Watershed Hydrology. Water. Resour. Publ., pp. 443–476.

Beven, K.J., 2005. Rainfall-runoff modelling: Introduction. In: Anderson, M.G. (Ed): Encyclopedia of Hydrological Sciences, Wiley, Chichester, pp. 1857–1868.

Brath, A., Montanari, A., Toth, E., 2004. Analysis of the effects of different scenarios of historical data availability on the calibration of a spatially-distributed hydrological model. Journal of Hydrology, 291, 3–4, 232–253. http://dx.doi.org/10.1016/j.jhydrol.2003.12.044.

Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A., 1984. Classification and Regression Trees. The Wadsworth and Brooks-Cole Statistics-Probability Series. Taylor & Francis, 368 p. ISBN: 0412048418, 9780412048418.

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

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

Chiew, F.H.S., Teng, J., Vaze, J., Post, D.A., Perraud, J.M., Kirono, D.G.C., Viney, N.R., 2009. Estimating climate change impact on runoff across southeast Australia: Method, results, and implications of the modeling method, Water Resour. Res., 45, W10414. DOI: 10.1029/2008WR007338.

Coron, L., Andréassian, V., Bourqui, M., Perrin, C., Hendrickx, F., 2011. Pathologies of hydrological model used in changing climatic conditions: a review. Hydro-climatology: Vari-

ability and change. In: Proceedings of IUGG2011 symposium J-H02, Melbourne, Australia.
Coron, L., Andréassian, 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. Water Resour. Res., 48, W05552. DOI: 10.1029/ 2011WR011721.
Coron, L., Andréassan, 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. DOI: 10.5194/hess-18-727-2014.
Das, T., Bárdossy, A., Zehe, E., He, Y., 2008. Comparison of conceptual model performance using different representations of spatial variability. J. Hydrol., 356, 106–118.
Farkas, C., Kværnø, S.H., Engebretsen, A., Barneveld, R., Deelstra, J., 2016. Applying profile and catchment-based mathematical models for evaluating the run-off from a Nordic catchment. J. Hydrol. Hydromech., 64, 3, 218–225. DOI: 10.1515/johh-2016-0022.
Fenicia, F., Kavetski, D., Savenije, H.H.G., 2011. Elements of a flexible approach for conceptual hydrological modeling: 1. Motivation and theoretical development. Water Resour. Res., 47, W11510. DOI: 10.1029/2010wr010174.
Finger, D., Heinrich, G., Gobiet, A., Bauder, A., 2012. Projections of future water resources and their uncertainty in a glacierized catchment in the Swiss Alps and the subsequent effects on hydropower production during the 21st century. Water Resour. Res., 48, 02521. DOI: 10.1029/2011WR010733, 2012.
Fowler, K.J.A., Peel, M.C., Western, A.W., Zhang, L., Peterson, T.J., 2016. Simulating runoff under changing climate conditions: Revising an apparent deficiency of conceptual rainfall-runoff models. Water. Resour. Res., 52, 1820–1846. DOI: 10.1002/2015WR018068.
Gaál, L., Szolgay, J., Kohnová, S., Parajka, J., Merz, R., Viglione, A., Blöschl, G., 2012. Flood timescales: Understanding the interplay of climate and catchment processes through comparative hydrology. Water Resour. Res., 48, W04511. DOI: 10.1029/ 2011WR011509.
Iorgulescu, I., Beven, K.J., 2004. Nonparametric direct mapping of rainfall-runoff relationships: An alternative approach to data analysis and modeling? Water Resour. Res., 40, W08403. DOI: 10.1029/2004WR003094.
Klemeš, V., 1986. Dilettantism in hydrology: Transition or destiny? Water Resour. Res., 22, 9, 177–188.
Kuentz, A., Arheimer, B., Hundecha, Y., Wagener, T., 2016. Understanding hydrologic variability across Europe through catchment classification. Hydrol. Earth Syst. Sci. Discuss., 21, 6, 1–28. DOI: 10.5194/hess-2016-428.
Magand, C., Ducharne, A., Le Moine, N., Brigode, P., 2015. Parameter transferability under changing climate: case study with a land surface model in the Durance watershed, France. Hydrological Sciences Journal, 60, 7–8, 1408–1423. DOI: 10.1080/02626667.2014.993643.
Merz, R., Blöschl, G., 2004. Regionalisation of catchment model parameters. Journal of Hydrology. 27, 95–123. DOI: 10.1002/hyp.6253.
Merz, R., Blöschl, G., Parajka, J., 2009. Scale effects in conceptual hydrological modelling. Water Resour. Res., 45, W09405. DOI: 10.1029/ 2009WR007872.
Merz, R., Parajka, J., Blöschl, G., 2011. Time stability of catchment model parameters: Implications for climate impact analyses. Water. Resour. Res., 47, 1015–1031. DOI: 10.1029/2010WR009505.
Nash, J.E. Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I-A discussion of principles, Journal of Hydrology, 10, 3, 282–290. DOI: 10.1016/0022-1694(70)90255-6.
Nester, T., Kirnbauer, R., Gutknecht, D., Blöschl, G., 2011. Climate and catchment controls on the performance of regional flood simulations. Journal of Hydrology, 340–356. http://dx.doi.org/10.1016/j.jhydrol.2011.03.028.
Nester, T., Komma, J., Blöschl, G., 2016. Real time forecasting in the Upper Danube basin. J. Hydrol. Hydromech., 64, 4, 404–414. DOI: 10.1515/johh-2016-0033.
Nijzink, R.C., Samaniego, L., Mai, J., Kumar, R., Thober, S., Zink, M., Schäfer, D., Savenije, H.H.G., Hrachowitz, M., 2016. The importance of topography-controlled sub-grid process heterogeneity and semi-quantitative prior constraints in distributed hydrological models. Hydrol. Earth Syst. Sci., 20, 1151–1176. DOI:10.5194/hess-20-1151-2016.
Osuch, M., Romanowicz, R.J., Booij, M.J., 2015. The influence of parametric uncertainty on the relationships between HBV model parameters and climatic characteristics. Hydrological Sciences Journal, 60, 7–8, 1299–1316. DOI: 10.1080/02626667.2014.967694.
Oudin, L., Perrin, C., Mathevet, T., Andréassian, V., and Michel, C., 2006. Impact of biased and randomly corrupted inputs on the efficiency and the parameters of watershed models. J. Hydrol., 320, 1–2, 62–83. DOI: 10.1016/j.jhydrol.2005.07.016.
Parajka, J., Blöschl, G., 2008. The value of MODIS snow cover data in validating and calibrating conceptual hydrologic models. Journal of Hydrology, 358, 3–4, 240–258. https://doi.org/10.1016/j.jhydrol.2008.06.006.
Parajka, J., Merz, R., Blöschl, G., 2005. A comparison of regionalisation methods for catchment model parameters. Hydrol. Earth Syst. Sci., 9, 157–171. DOI: 10.5194/hess-9-157-2005.
Parajka, J., Merz, R., Blöschl, G., 2007. Uncertainty and multiple calibration in regional water balance modelling case study in 320 Austrian catchments. Hydrol. Process, 21, 435–446. DOI: 10.1002/hyp.6253.
Pechlivanidis, I.G., Arheimer, B., 2015. Large-scale hydrological modelling by using modified PUB recommendations: the India-HYPE case. Hydrol. Earth Syst. Sci., 19, 4559–4579. DOI: 10.5194/hess-19-4559-2015.
Pebesma, E.J., 2001. Gstat User's Manual. Dep. of Phys. Geogr., Utrecht Univ., Utrecht, The Netherlands.
Perrin, C., Michel, C., Andréassian, V., 2001. Does a large number of parameters enhance model performance? Comparative assessment of common catchment model structures on 429 catchments. J. Hydrol., 242, 275–301. https://doi.org/10.1016/S0022-1694(00)00393-0.
Perrin, C., Michel, C., Andréassian, V., 2003. Improvement of a parsimonious model for streamflow simulation. J. Hydrol., 279, 275–289. DOI: 10.1016/s0022-1694(03)00225-7.
Perrin, C., Oudin, L., Andréassian, V., Rojas-Serna, C., Michel, C., Mathevet, T., 2007. Impact of limited streamflow data on the efficiency and the parameters of rainfall-runoff models. Hydrol. Sci. J., 52, 1, 131. http://dx.doi.org/10.1623/hysj.52.1.131.
Perrin, C., Andréassian, V., Rojas-Serna, C., Mathevet, T., Le Moine, N., 2008. Discrete parameterization of hydrological models: Evaluating the use of parameter sets libraries over 900 catchments. Water Resour. Res., 44, W08447. DOI: 10.1029/2007WR006579.

Poncelet, C., Merz, R., Parajka, J., Oudin, L., Andréassian, V., Perrin, C., 2017. Process-based interpretation of conceptual hydrological model performance using a multinational catchment set. Water Resource Research. DOI: 10.1002/2016WR019991.

R Development Core Team, 2011. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Saft, M., Western, A.W., Zhang, L., Peel, M.C., Potter, N.J., 2015. The influence of multiyear drought on the annual rainfall-runoff relationship: An Australian perspective. Water Resour. Res., 51, 2444–2463. DOI: 10.1002/2014WR015348.

Saft, M., Peel, M.C., Western, A.W., Zhang, L., 2016. Predicting shifts in rainfall-runoff partitioning during multiyear drought: Roles of dry period and catchment characteristics. Water Resour. Res., 52. DOI: 10.1002/ 2016WR019525.

Schaefli, B. Gupta, H.V., 2007. Do Nash values have value? Hydrol. Process., 21, 2075–2080. DOI: 10.1002/hyp.6825.

Seibert, J., 2003. Reliability of model predictions outside calibration conditions. Nordic Hydrology, 34, 477–492.

Seibert, M., Merz, B., Apel, H., 2016. Seasonal forecasting of hydrological drought in the Limpopo basin: A comparison of statistical methods. Hydrol. Earth Syst. Sci. Discuss. DOI: 10.5194/hess-2016-4, 2016.

Seifert, D., Sonnenborg, T.O., Refsgaard, J.C., Højberg, A.L., Troldborg, L., 2012. Assessment of hydrological model predictive ability given multiple conceptual geological models. Water Resour. Res., 48, W06503. DOI: 10.1029/2011WR011149.

Seiler, G., Anctil, F., Perrin, C., 2012. Multimodel evaluation of twenty lumped hydrological models under contrasted climate conditions. Hydrol. Earth Syst. Sci., 16, 4, 1171–1189. http://dx.doi.org/10.5194/hess-16-1171-2012.

Sleziak, P., Szolgay, J., Hlavčová, K., Parajka, J., 2016a. The impact of the variability of precipitation and temperatures on the efficiency of a conceptual rainfall-runoff model. Slovak Journal of Civil Engineering, 24, 4, 1–7. DOI: 10.1515/sjce-2016-0016.

Sleziak, P., Szolgay, J., Hlavčová, K., Parajka, J., 2016b. Assessment of the performance of a hydrological model in relation to selected climatic characteristics. In: Proc. 16th International Multidisciplinary Scientific GeoConference SGEM 2016, Book 3 Vol. 3, pp. 43–52. DOI: 10.5593/SGEM2016/HB33/S02.006.

Stauer, J.J., Stensvold, K.A., Gregory, M.B., 2010. Determination of biologically significant hydrologic condition metrics in urbanizing watersheds: an empirical analysis over a range of environmental settings. Hydrobiologia, 654, 1, 27–55. DOI: 10.1007/s10750-010-0362-0.

Sun, W., Wang, Y., Wang, G., Cui, X., Yu, J., Zuo, D., Xu, Z., 2017. Physically based distributed hydrological model calibration based on a short period of streamflow data: case studies in four Chinese basins. Hydrol. Earth Syst. Sci., 21, 251–265. DOI: 10.5194/hess-21-251-2017.

Therneau, T., Atkinson, B., Ripley, B., 2017. Recursive partitioning and regression trees. Version 4.1-11.

van Esse, W.R., Perrin, C., Booij, M.J., Augustijn, D.C.M., Fenicia, F., Kavetski, D., Lobligeois, F., 2013. The influence of conceptual model structure on model performance: a comparative study from 273 French catchments. Hydrol. Earth Syst. Sci., 17, 4227–4239. DOI: 10.5194/hess-17-4227-2013.

van Griensven, A., Meixner, T., Grunwald, S., Bishop, T., Diluzio, M., Srinivasan, R., 2006. A global sensitivity analysis tool for the parameters of multi-variable catchment models. Journal of Hydrology, 324, 10–23.

Valent, P., Szolgay, J., 2012. Assessment of the uncertainties of a conceptual hydrologic model by using artificially generated flows. Slovak Journal of Civil Engineering, 20, 4, 35–43. DOI: https://doi.org/10.2478/v10189-012-0020-9.

Vaze, J., Post, D.A., Chiew, F.H.S., Perraud, J.M., Viney, N.R., Teng, J., 2010. Climate nonstationarity – Validity of calibrated rainfall-runoff models for use in climatic changes studies. J. Hydrol., 394, 3–4, 447–457. DOI: 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. DOI: 10.5194/hess-17-2263-2013.

Viglione, A., Parajka, J., 2014. TUWmodel: Lumped hydrological model for educational purposes. Version 0.1-4. https://cran.r-project.org/web/packages/TUWmodel/index.html.

Viviroli, D., Zappa, M., Schwanbeck, J., Gurtz, J., Weingartner, R., 2009. Continuous simulation for flood estimation in ungauged mesoscale catchments of Switzerland – Part I: Modelling framework and calibration results. Journal of Hydrology, 377, 191–207. https://doi.org/10.1016/j.jhydrol. 2009. 08.023.

Wang-Erlandsson, L., Bastiaanssen, W.G.M., Gao, H., Jagermeyer, J., Senay, G.B., van Dijk, A.I.J.M., Guerschman, J.P., Keys, P.W., Gordon, L.J., Savenije, H.H.G., 2016. Global root zone sorage capacity from satellite-based evaporation. Hydrol. Earth Syst. Sci., 20, 1459–1481. www.hydrol-earth-syst-sci.net/20/1459/2016/.

Wilby, R.L., 2005. Uncertainty in water resource model parameters used for climate change impact assessment. Hydrol. Processes, 19, 16, 3201–3219.

Received 6 October 2017
Accepted 23 April 2018