TESTING OF AN ALTERNATIVE APPROACH TO CALIBRATION OF A HYDROLOGICAL MODEL UNDER VARYING CLIMATIC CONDITIONS

Patrik Sleziak, Michal Danko, Ladislav Holko

Conceptual rainfall-runoff models are routinely used in practical water resource investigations. Common uncertainties associated with these models (in addition to the uncertainty related to schematization and structure of the models) include for example errors in the inputs, calibration/validation uncertainties (e.g., choice of the suitable lengths of the two periods), uncertainties related to the use of the models in other climatic conditions, etc. This study addresses the uncertainties related to the choice of calibration/validation periods for the long data sets with varying climatic inputs. It is conducted in the pilot catchment of the Jalovecký Creek (area 22.2 km$^2$) in Slovakia and uses data from the 30-years long period 1989–2018. A HBV type model (the IUW model) is used for the modelling. Two different approaches to selection of calibration period are compared. In the first approach, the calibration period is determined by division of the available data into three equally long periods (each of them is then used in model calibration and validation). Such an arbitrary division is the common practice in hydrological modelling. In the second approach, the selection of calibration periods is based on the cycles found in the measured data. The wavelet transform method revealed cyclical components in air temperature with period of 6-years. Periods in other data sets were less significant. In accordance with this finding, the model is calibrated for five 6-years long periods. Model performance for the two approaches to selection of the calibration periods is evaluated by visual comparison of measured and simulated monthly flows in different climatic periods and by the Nash-Sutcliffe efficiency coefficient. The two approaches to the selection of calibration period provided similar results. However, the model calibrated in colder period represents monthly flows more reliably than the model that was calibrated in warmer period. In terms of predictions related to climate change impacts it would mean that hydrological models calibrated in current period should provide reasonable simulations for warmer climate.

KEY WORDS: hydrological model calibration, method of wavelet transform, climate change

Introduction

Conceptual rainfall-runoff (r-r) models are routinely used in various practical water resource investigations (for example, flow forecasting, flood impact assessment, climate impact studies, etc.). Common problem linked with the use of these models is their calibration (i.e. the way to deal with uncertainties in parameter estimation, choice of the suitable length of calibration and validation periods, etc.). Previous studies (e.g., Merz et al., 2009; Perrin et al., 2007) showed that the length of the calibration period may significantly affect model calibration and model performance (for example, if the calibration period is too short). Merz et al. (2009) calibrated a semi-distributed conceptual r-r model for periods of 1, 3, 5, 10, 15, 30 years and analyzed the effect of the length of the calibration period on model performance. Their study was performed for 269 catchments in Austria. They found that the model performance during calibration period decrease and the model performance during validation period increase with the number of years available for calibration. Their results suggest that minimum calibration period to achieve good model performance is five years. Other authors suggest that the optimal length for the model calibration may vary from two to ten years (Anctil et al., 2004; Brath et al., 2004). Generally, the calibration period should be long enough to capture the variability of climatic and flow conditions (Sorooshian et al., 1983).

Previous studies (e.g., Vaze et al., 2010; Merz et al., 2011; Coron et al., 2012; Fowler et al., 2016; Saft et al., 2016; Sleziak et al., 2017, 2018) demonstrated that the r-r models show significant reductions in performance when used in climatic conditions that differ from the conditions for which were the models calibrated. For example, in Australia, Vaze et al. (2010) and Coron et al. (2012) observed that the hydrological model had a tendency to overestimate mean runoff when the calibration period was wetter (i.e., a wet to dry parameter transfer). On the contrary, in Austria, Merz et al. (2011) and Sleziak et al. (2018) showed that a HBV model calibrated in a colder/drier decade (e.g., 1981–1990)
tended to overestimate the runoff in a warmer and wetter decade (e.g., 2001–2010), particularly in flatland basins. Such findings naturally indicate problems with model schematization or structure. In theory, the performance of a model that captures the dominant hydrological processes correctly, should not depend too much on the climatic (i.e. input) data. However, changing climate may result in a change of the dominant processes. Numerous studies were carried out on this topic, but the results are not always consistent among them due to different regions, physiographic conditions, etc. Proper approaches in hydrological modelling under changing climate therefore remains a great challenge for the modelers.

The objective of this article is to test an alternative approach to calibration of a lumped HBV model for the use in changing (or changed) climate. Specifically, we asked two research questions: (1) Can the determination of calibration periods based on natural cycles defined by the data result in better simulation results? (2) Can hydrological models that are calibrated in current climatic conditions achieve satisfactorily results in the future warmer climate?

The paper is organized as follows. The data section describes the study area and data. The methods section gives the details on determination of natural cycles in the data series, the TUW model and its application. The results section summarizes the results of this study. The discussion section discuss and compares the results to other studies. The last section presents conclusions.

### Study area and data

This study is carried out in the mountain catchment of the Jalovecký Creek in Slovakia (Fig. 1). The catchment is representative for the hydrological cycle of the highest part of the Western Carpathians. Catchment area is 22.2 km². Elevations in the catchment range from 800 to 2178 m a.s.l. (mean 1500). Mean slope is 30°. Soil cover is represented by Cambisols, Podsol, Ranker and Litho-soils. Forests dominated by spruce cover 44% of catchment area, dwarf pine covers 31% and alpine meadows and bare rocks cover the rest 25%.

Daily catchment precipitation daily air temperature, discharge and potential evapotranspiration from the period 1989–2018 are used. Point precipitation and air temperature measurements were interpolated to obtain catchment values. Potential evapotranspiration was calculated by the Blaney-Criddle method (Schröder, 1985).

### Methods

Two approaches to division of input data for the model (precipitation and air temperature as climatic drivers and runoff used for the validation of simulation) are used. First, the data are divided arbitrarily into equal periods, i.e. decades (1989–1998, 1999–2008, 2009–2018). Second, the data are divided into periods identified by the analysis of data carried out by the wavelet transform method. The idea is to use the data from the same natural cycle in hydrological modelling.

#### Method of wavelet transform

It is well know that the natural processes occur in cycles (e.g., Hurst, 1951; Klement 1974; Pekárová and Pekár, 2007). The WT method is a simple method to estimate the changes in cyclical components and variability of the time series (e.g., Sabo 2012). We apply the method to daily time series of air temperature, precipitation, and discharges of the Jalovecký Creek catchment. The R software environment (R Development Core Team, 

![Fig. 1. The location and topography of the Jalovecký Creek catchment.](image_url)
and the package WaveletCo (Tian and Cazelles, 2011) are used to conduct the analysis that consists of four steps. First, the database of hydrometeorological time series (daily air temperature, precipitation, discharges) is created. Then, the work package is started and the models for the descriptions of changes in the cyclical components and the variability of the time series are created. In the last step, the structural changes in hydrometeorological time are analyzed/interpreted. For the interpretation of the results the scalogram used (see the Results).

**Modelling**

We apply the simple lumped model TUW (Viglione and Parajka, 2014), which follows the structure of a widely used Swedish HBV model (Bergström, 1995). The model simulates daily discharge using daily precipitation, air temperature, and potential evapotranspiration as inputs. It has 15 parameters that need to be calibrated. The model involves three modules (i.e., snow, soil, runoff module). In the snow module, the accumulation and melt are computed by the degree-day method. Groundwater recharge and actual evaporation are functions of actual water storage in the soil module. In runoff module, the runoff formation is represented by two linear reservoir equations. Channel routing is simulated by a triangular weighting function. More details about the model are given in Parajka et al. (2007).

The model is automatically calibrated using a differential evolution algorithm Deoptim (Ardia et al., 2015). The objective function consists of the combination of the Nash-Sutcliffe coefficient (NSE, Nash and Sutcliffe, 1970) and the logarithmic Nash-Sutcliffe coefficient (logNSE, Merz et al., 2011). The function is used in the form of \((\text{NSE} + \log\text{NSE})/2\).

Calibration and validation periods determined by the two approaches, i.e. the arbitrary and cycles-based divisions are used in the modelling. The model is in both cases calibrated and validated consecutively for each period and the comparison of results represent the Differential Split-Sample Test (DSST) proposed by Klemes (1986) for model performance testing. Because the climatic characteristics in individual periods differ, the DSST evaluates model performance in periods with contrasting climate. The commonly used Nash–Sutcliffe Efficiency (NSE) is used to evaluate the model performance. A good simulation result will have NSE close to 1 (the best match between observed and simulated flows).

**Results**

**Identification of cyclical components**

Significant cyclical component was determined only for the air temperature data. The result is presented in the form of scalogram in Fig. 2. The scalogram provides an information about two parameters: scale (i.e., frequency) and time. The horizontal axis denotes time (i.e., duration of the signal) and the vertical axis denotes period (i.e., cyclical components). The size of the wavelet coefficients (used to receive an estimation of power spectrum) can be identified by the intensity of color. Significant periods are identified by the white line drawn around the intense colors. The section of the scalogram outside the cupola shape is considered problematic. The right side of the scalogram shows the global wavelet spectrum (averaged scalogram). The scalogram shown in Fig. 2 identifies significant periods in a range of 6–8 years, which occurred between the years 1995–2010. Because scalograms for precipitation and discharge did not show any significant cycles, the 6 years long period obtained from the air temperature data was used to split the data for the modelling into five periods.

Comparison of air temperature, precipitation and discharge in the periods determined arbitrarily and according to cycles identified in the data is shown in Fig. 3. It is clear that the period 1999–2008 is colder. The values of the mean annual air temperatures decrease on average by 0.7°C between periods 1989–1998 and 1999–2008 and 0.4°C between periods 1999–2008 and 2009–2018. Precipitation shows a slightly decreasing trend over the entire study period while the mean annual discharge does not change. The right part of Fig. 3 shows that air temperature has decreasing trend approximately until the year of 2006 with decrease of about 1.5°C between

![Fig. 2. Scalogram of air temperatures at Jalovecký Creek catchment. The data indicate significant cycles with period of 6-8 years that occurred approximately in 1995–2010.](image-url)

Assessment of the model performance for two approaches to selection of calibration/validation periods

Results of model calibration and validation (in terms of different calibration lengths) are summarized in Tables 1 and 2. The results show that the model performed better in calibration periods. For the first calibration approach (i.e., the calibration periods are determined arbitrarily, Table 1), the NSE values in calibrations are on average 0.67 (1989–1998), 0.75 (1999–2008), and 0.76 (2009–2018). For the second calibration approach (i.e., determination of calibration periods based on naturally cycles, Table 2), the NSE values in calibrations are on average 0.63 (1989–1994), 0.75 (1995–2000), 0.73 (2001–2006), 0.74 (2007–2012) and 0.80 (2013–2018). The NSE values in the validation periods are generally lower than in the calibration periods. Different length or periods obtained by the two approaches did not result in significant differences.

Performances in the colder calibration period (1999–2008) are satisfactory, with NSE of 0.75. The NSE in the warmer validation period is 0.73 (Table 1). Similar result in terms of NSE is also achieved in the calibration period 2007–2012 (Table 2). Generally, the model calibrated in colder decade gives more satisfactorily results. This is also demonstrated in comparison of simulated and measured monthly flows (Figures 4 and 5). For the simulations we used parameters from colder calibration periods (i.e., period 1999–2008 and period 2007–2012). Comparison of measured and simulated monthly flows indicates that the model provided reasonable simulation of flows in different calibration and validation periods. Our results suggest that the model calibrated in colder decade more reliably represent the monthly flows. This implies that hydrological models calibrated in current climate (2013–2018) could provide reasonable predictions for the changed climate (i.e., for future warmer conditions).
Table 1. Values of the Nash-Sutcliffe coefficient (NSE) in calibration and validation periods with length of 10 years; shaded area indicates the NSE values obtained in the calibrations; WP is warmer period and CP is colder period.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0.67</td>
<td>0.70</td>
<td>0.63</td>
<td></td>
</tr>
<tr>
<td>0.59</td>
<td>0.75</td>
<td>0.73</td>
<td></td>
</tr>
<tr>
<td>0.48</td>
<td>0.68</td>
<td>0.76</td>
<td></td>
</tr>
</tbody>
</table>

Discussion

We evaluated the influence of selection of calibration and validation periods on performance of a lumped hydrological model. Length of the data series used in calibration and validation was 10 years and 6 years, respectively. Visual examination and the model performance efficiency indicator (NSE) show that the differences are small. Several studies (e.g., Merz et al. 2009; Perrin et al., 2007) addressed the question how much data are needed for model calibration. Merz et al. (2009) in their study showed that the calibration period of five years would be the minimum for achieving a good model performance. Other results suggest that the optimal length for the model calibration may vary from two to ten years (Anctil et al., 2004; Brath et al., 2004). Our results indicate that the length of calibration periods 6 and 10 years are both appropriate achieving similar simulations. It suggests that if the calibration period is long enough to cover periods of natural cycles in the input data, the arbitrary splitting of data into calibration and validation periods does not produce worse simulations. This conclusion should be verified in other catchments.

Studies show that the r-r models provide poor simulations when are applied in changing climate (e.g., Merz et al., 2011; Vaze et al., 2010; Coron et al., 2012; Sleziak et al., 2018). Merz et al. (2011) found that the model calibrated
Table 2. Values of the Nash-Sutcliffe coefficient (NSE) and in calibration and validation periods with length of 6 years; shaded area indicates the NSE values obtained in the calibrations; WP is warmer period and CP is colder period

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>NSE</td>
<td>0.63</td>
<td>0.69</td>
<td>0.62</td>
<td>0.65</td>
<td>0.58</td>
</tr>
<tr>
<td></td>
<td>0.57</td>
<td>0.75</td>
<td>0.61</td>
<td>0.62</td>
<td>0.56</td>
</tr>
<tr>
<td></td>
<td>0.56</td>
<td>0.65</td>
<td>0.73</td>
<td>0.65</td>
<td>0.65</td>
</tr>
<tr>
<td></td>
<td>0.51</td>
<td>0.64</td>
<td>0.63</td>
<td>0.74</td>
<td>0.73</td>
</tr>
<tr>
<td></td>
<td>0.23</td>
<td>0.34</td>
<td>0.39</td>
<td>0.60</td>
<td>0.80</td>
</tr>
</tbody>
</table>

Fig. 5. Comparison of the observed (blue colour) and simulated (red colour) monthly flows for simulations based on the cycles-based division of the data (i.e., into five 6-years long periods). The simulations were performed with the parameters obtained in the calibration period of 2007–2012.
in a colder time period had tendency to overestimate the runoff (Q95 by about 12%, Q50 by about 15%, Q5 by about 35%) in a warmer/wetter period. Sleziak et al. (2018) observed overestimation of mean runoff when the calibration period was colder too. Coron et al. (2012) and Vaze et al. (2010) for Australian catchments observed a tendency to overestimate runoff when the calibration period was wetter. According to the results, the best Nash-Sutcliffe efficiency was achieved in calibration period of 2013–2018 (warmer period) (Table 2), but validation in years 1989–1994, 1995–2000, 2001–2006 showed worse results compared with those from the first selection of calibration periods (Table 1). This can be related to (a) different length of calibration periods, (b) algorithm used in the calibration strategy. Underestimation of measured flows in spring months (April and May) indicate the need to think about the improvement of the snowmelt-rainfall phase of runoff formation.

Conclusions

In this study, we have assessed the uncertainties related to the choice of calibration/validation periods. For the modelling purposes, we used a popular HBV model. From our results the following conclusions can be drawn. The model’s uncertainties are associated with the climatic characteristics. In other words, the difference in climatic characteristics between the calibration and validation periods affects the performance of hydrological model. We found that the model calibrated in colder periods (i.e., 1999–2008 and 2007–2012) provided better representation of flows. This finding point to the fact that hydrological models calibrated in current climatic conditions could work reasonably also in the future warmer climate. These results should be validated in other catchments. In the future we plan (a) to extend such analysis to other regions (e.g., Slovakia vs. Austria), (b) to include other data (e.g., soil moisture data, MODIS data, etc.) to calibration strategy and thus better represent the hydrological components, (c) to use HBV vs. spatially distributed based models.

Acknowledgement

This study was supported by the grants from the Slovak Academy of Sciences (project VEGA No. 2/0065/19) and from the Slovak Research and Development Agency (project APVV No. 15-0497).

References

Coron, L., Andréassian, V., Perrin, C., Lerat, J., Vaze, J., Bourqui, M., Hendrickx, F. (2012): Crash testing...


