The role of conceptual hydrologic model calibration in climate change impact on water resources assessment
Andrijana Todorovic and Jasna Plavsic

ABSTRACT
Assessment of climate change (CC) impact on hydrologic regime requires a calibrated rainfall-runoff model, defined by its structure and parameters. The parameter values depend, inter alia, on the calibration period. This paper investigates influence of the calibration period on parameter values, model efficiency and streamflow projections under CC. To this end, a conceptual HBV-light model of the Kolubara River catchment in Serbia is calibrated against flows observed within 5 consecutive wettest, driest, warmest and coldest years and in the complete record period. The optimised parameters reveal high sensitivity towards calibration period. Hydrologic projections under climate change are developed by employing (1) five hydrologic models with outputs of one GCM–RCM chain (Global and Regional Climate Models) and (2) one hydrologic model with five GCM–RCM outputs. Sign and magnitude of change in projected variables, compared to the corresponding values simulated over the baseline period, vary with the hydrologic model used. This variability is comparable in magnitude to variability stemming from climate models. Models calibrated over periods with similar precipitation as the projected ones may result in less uncertain projections, while warmer climate is not expected to contribute to the uncertainty in flow projections. Simulations over prolonged dry periods are expected to be uncertain.

Key words | climate change, hydrologic modelling, Kolubara River, parameter transferability

INTRODUCTION
The climate change (CC) impact on a hydrologic regime is commonly assessed using the outputs of a Global Climate Model (GCM) as an input to a calibrated hydrologic model (e.g. Wilby & Harris 2006; Graham et al. 2007; Kay et al. 2009; Chiew et al. 2010; Najafi et al. 2011; Teng et al. 2012). Because of their coarse spatial resolution, GCM outputs have to be downscaled to be suitable for hydrologic modelling. Once calibrated, hydrologic models are assumed capable of simulating future runoff. Simulated hydrologic variables for future are compared to those simulated in a reference historical period, so that changes in a hydrologic regime can be estimated. This approach, however, is fraught with uncertainties arising from the assumed greenhouse gas (GHG) emission scenario, GCM, downscaling method, hydrologic model structure and parameter estimates (Blöschl & Montanari 2010).

There are numerous studies aiming at analysing these uncertainties. For example, Kay et al. (2009) studied uncertainties due to emission scenario, GCMs, downscaling method, hydrologic model structure and parameter estimates. Their results indicated GCMs to be the primary source of uncertainty. Wilby & Harris (2006), Najafi et al. (2011) and Teng et al. (2012) also indicated GCMs as most significant source of uncertainty. Ability of GCMs to reproduce observed rainfall patterns is also questioned in the literature. Vaze et al. (2011) found that GCMs reproduced spatial distribution and mean precipitation, but their ability to replicate the entire time series was limited. For a CC impact study, Tan et al. (2014) used only the most skilful GCMs in terms of their ability to reproduce historical precipitation. Kay et al. (2009) showed that different downscaling methods result in different precipitation depths in the historical
period. In addition, different GCM downscaling methods yield quite divergent river flow projections (Chiew et al. 2010).

Uncertainties resulting from hydrological models have generally been shown to be smaller compared to those due to GCMs or gas emission scenarios, though they may be considerable (Wilby 2005; Bae et al. 2011; Bastola et al. 2011). Najafi et al. (2011) found that the uncertainties coming from the hydrologic model exceed those due to GCM over dry seasons. Jiang et al. (2007) and Bastola et al. (2011) argued that uncertainties in flow projections stemming from hydrological models have not been sufficiently investigated, and that further research in this domain is needed. Blöschl & Montanari (2010) hypothesised that better understanding of complex interactions between climate and hydrological processes under current climate would lead to reduction in uncertainties in hydrologic projections.

Hydrologic model structure has an impact on variability of river flow projections. Jiang et al. (2007) and Bae et al. (2011) demonstrated that different hydrologic models, given the same climate forcing, project different seasonal distributions of runoff, evapotranspiration (ET) and soil moisture, despite similar performance over the observation period. In addition, uncertainties in flow projections also stem from selection of the method for estimating potential evapotranspiration (PET). The results of Bae et al. (2011) revealed that different PET methods introduce additional variance in the projections, which is shown to be larger in distant future because of greater differences in temperature. The uncertainties arising from hydrologic model structure and parameter assessment and those from GHG emission scenario and GCMs were compared by Bastola et al. (2011), who calibrated four conceptual hydrologic models within the Generalised Likelihood Uncertainty Estimation (GLUE) framework (Beven & Binly 1992). Their results demonstrated that hydrologic models cause considerable uncertainty.

The values of the hydrologic model parameters are typically inferred through the calibration procedure, in which the parameter values are adjusted to achieve a satisfactory agreement between the observed and simulated state variables according to some goodness-of-fit measures (Beven 2001; Yilmaz et al. 2010). Values of the parameters depend on the objective function(s) (Yapo et al. 1996), optimisation method, data errors (i.e. measurement errors and inadequate data resolution) and properties of the calibration period (Deletic et al. 2012). Uncertainty in hydrologic parameters reflects the hydrologic projections. For example, Najafi et al. (2011) showed that the models calibrated using different objective functions resulted in similar projected changes in winter runoff, but the summer runoff projections varied to a considerable extent.

Model calibration over different periods has been proven to result in different optimal parameter sets, even if the same objective functions and optimisation methods are applied (Klemeš 1986; Gharari et al. 2013). For example, Wagener et al. (2003) calibrated a lumped conceptual model in the moving time windows applying the GLUE method, and demonstrated that the posterior distributions of all parameters change in time. Merz et al. (2011) examined for temporal trends in the HBV model parameters by calibrating the model in consecutive 5-year periods at 273 catchments, and identified consistent trends in some snow- and soil-related parameters. Osuch et al. (2014) expanded this research by quantifying correlation between the optimised HBV parameters and numerous climatic indices. They detected statistically significant correlation between maximum soil storage and the PET- and temperature-related indices. Choi & Beven (2007), de Vos et al. (2010) and Zhang et al. (2011) calibrated hydrologic models over different periods defined by centroid-based clustering of climate variables (e.g. precipitation, PET, aridity indices, etc.). All these studies demonstrated that the parameters’ distributions vary over different periods, i.e. clusters.

Since the optimal parameter values change with the calibration period, a perfectly consistent performance of a hydrologic model cannot be obtained (Abebe et al. 2010; Gharari et al. 2013). Consistency in model performance is assessed with the Split Sample Test (SST) or Differential Split Sample Test (DSST) (Klemeš 1986). The former implies model evaluation in an independent but similar period, while the latter implies quite a different evaluation period and is considered more robust (Klemeš 1986; Seibert 2003; Coron et al. 2012). Many authors performed DSST on periods that differ in terms of meteorological properties (e.g. precipitation depth) or in terms of the calibration period length (Seibert 2003; Vaze et al. 2010; Li et al. 2012; Luo et al. 2012). Their results revealed a substantial decrease
in model performance in the evaluation periods, thereby proving that the choice of the calibration period is crucial for model performance. Further, application of DSST to the entire modelling chain (emission scenario, GCM, Regional Climate Model (RCM) and hydrologic model) was recommended by Refsgaard et al. (2014). Recommendations on DSST application can be found in Thirel et al. (2014).

Dependence of optimal parameter sets on the calibration period may generate additional uncertainty in hydrologic projections under CC. This uncertainty may increase with the time lag between calibration and simulation periods (Merz et al. 2011). However, there is little research on uncertainties in hydrologic projections due to selection of the calibration period. Wilby (2005) calibrated two versions of a semi-distributed conceptual model in the wettest year, the driest year, in a year that was analogous to the projected conditions in 2050’s, and over 1961–1990. Simpler version produced significant differences (up to 40%) among flow predictions for different calibration periods. These differences were comparable in magnitude to those due to an emission scenario. The complex version of the model resulted in similar flow predictions, regardless of the calibration period. However, the robustness of a model calibrated over 1 year is questionable. Brigode et al. (2013) calibrated two lumped, conceptual models at 89 catchments in the full record period, and in the wettest, intermediate and dry 3-year periods. They demonstrated that the flow projections could be rather different if calibration is performed in different periods, either when a single optimal parameter set or an ensemble of parameter sets is used for hydrologic simulation. Magand et al. (2014) calibrated a semi-distributed Catchment Land Surface Model (CLSM) in a multi-objective manner over consecutive 9-year periods and in the full record period. They used one Pareto-optimal parameter from each period to obtain hydrologic projections. Although these parameter sets produced similar results with DSST, projected future flows and ET rates were quite different.

The available studies indicate that the choice of the calibration period for a hydrologic model is an important issue in CC impact studies, since the climate characteristics of these periods are crucial for model performance. Having in mind the projected CCs, especially temperature increase, questions remain about which calibration strategy would be the best for CC impact studies. To our knowledge, differences in hydrologic projections obtained by employing models calibrated in contrasted periods in terms of temperature have not been analysed so far. The aim of this paper is to investigate further the impact of calibration period on hydrologic model parameters, model efficiency and hydrologic projections under changing climate. To this end, a conceptual HBV-light rainfall–runoff model of the Kolubara River catchment in Serbia is calibrated in four contrasted periods in terms of precipitation depths and temperature, and in the full hydrologic record period. Every model version is evaluated over the remaining four periods, and employed to obtain hydrologic projections in two future periods using the results of a single GCM–RCM chain. These projections are compared to the projections obtained with a calibrated hydrologic model and outputs of five different GCM–RCM chains. In this way, assessment of (1) model transferability between different periods (especially in terms of temperature, which is particularly important in CC impact studies), and (2) the impact of calibration period on hydrologic projections is enabled. In accordance with the results, recommendations for hydrologic model application in CC impact studies and further research are given.

DATA AND METHODS

Catchment

The Kolubara River is a 123 km long right tributary of the Sava River, with the total catchment area of 3,638 km² (Figure 1). Arable cultivated land and deciduous forests prevail in the catchment (over 80% of the total area), while urban areas constitute approximately 2% (Nestorov & Protić 2009). Forests dominate the area above 600 m above sea level (a.s.l.) (Langsholt et al. 2013).

The Kolubara River has a mixed rainfall-snowmelt water regime, with considerable flow fluctuations (Langsholt et al. 2013). The highest flows occur in March and April, while the lowest flows are observed in August and September. Mean annual precipitation observed at the Valjevo meteorological station (176 m a.s.l.) and estimated ET for 1954–2010 amounts to 790 mm/year and 481 mm/year, respectively. The linear regression slope test revealed an increase in
mean annual temperatures, and an absence of a statistically significant trend in precipitation and flows.

The Slovac stream gauge is situated 88 km upstream of the confluence of the Kolubara and Sava rivers. The drainage area to the Slovac stream gauge is 995 km$^2$ (hatched area in Figure 1), with elevations ranging from 122 to 1,346 m a.s.l. Mean annual runoff at Slovac for 1954–2010 amounts to 308.8 mm/year, i.e. mean flow is 9.7 m$^3$/s.

**Data**

Daily precipitation depths and mean daily temperatures observed at the Valjevo meteorological station from 1949 to 2010, and mean daily flows observed at Slovac from 1954 to 2010 were made available for model application from the Republic Hydrometeorological Service of Serbia (RHMSS).

Monthly PET rates, calculated using the Eagleman’s method (Eagleman 1967; Bergström 1976; Bergström et al. 1992) and it has commonly been used for stream flow simulations, forecasting and CC impact assessment (Xu 2000; Graham et al. 2007; Langsholt et al. 2013). The HBV-light is based on the HBV model: it consists of the snow, soil and response routines, and it is enhanced by the modules for automatic calibration and batch simulation (Seibert & Vis 2012).

**Snow routine**

Precipitation at temperature below the threshold temperature TT is considered snow. Owing to underestimation of measured snow and snowpack evaporation, a snow correction factor (SF$_{CI}$) is introduced. Snowmelt is modelled with the degree-day method: melting is proportional (CF$_{max}$ parameter) to the difference between air temperature and TT. Water refreezing and meltwater retainment by the snowpack are also modelled. Precipitation and temperature are corrected to account for changes with elevation (parameters PCALT and TCALT) (Seibert & Vis 2012).

**Soil routine**

Soil is represented by two reservoirs (Figure 2): upper (SUZ), with maximum capacity of FC, and lower zone
box (SLZ). The amount of precipitation and snowmelt to be retained in SUZ reservoir, and the amount to percolate to SLZ (less or equal to the PERC parameter value) are calculated using a non-linear function (parameter $\beta$) of SUZ water content. Evaporation from SUZ depends on the water content and occurs at a potential rate if it exceeds $L_p \cdot FC$, where $L_p$ defines reduction in ET (Seibert & Vis 2012).

**Response function**

Discharges from the reservoirs are calculated using linear outflow equations. The sum of the outflows is transformed using a triangular function (parameter MAXBAS), resulting in total runoff.

**Input data**

HBV-light requires daily precipitation, mean daily temperature and observed mean daily runoff. PET is represented as daily rate for every month of a representative year.

**Model calibration**

The HBV-light model can be lumped or semi-distributed in terms of vegetation and/or elevation zones. Parameters can be calibrated using the built-in GAP algorithm (Genetic Algorithm and Powell’s method). Model calibration begins with random selection of parameter values from the prior uniform or normal distribution, to establish an initial parameter set. An optimised parameter set is generated by employing genetic algorithm, followed by application of the Powell’s method for local optimisation (Seibert 2000; Seibert & Vis 2010, 2012). Model parameters and their initial ranges (prior uniform distribution) adopted in this paper are given in Table 1.

**Model calibration and evaluation**

To estimate significance of the calibration period, HBV-light model is calibrated in the full hydrologic record period.

---

**Table 1** | HBV-light parameters: initial ranges and optimised values

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Snow</td>
<td>$T_T$ [C]</td>
<td>-2.5</td>
<td>2.5</td>
<td>-1.5</td>
<td>0.2</td>
<td>0</td>
<td>-1.9</td>
<td>-0.75</td>
</tr>
<tr>
<td></td>
<td>$C_{F_{max}}$ [mm/day/ C]</td>
<td>0.1</td>
<td>10</td>
<td>4.9</td>
<td>3.5</td>
<td>4.9</td>
<td>3.9</td>
<td>4.8</td>
</tr>
<tr>
<td></td>
<td>$S_{FCF}$ [-]</td>
<td>0.5</td>
<td>1.5</td>
<td>1.3</td>
<td>1.2</td>
<td>0.8</td>
<td>1.1</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td>$C_{FR}$ [-]</td>
<td>0</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.05</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>$C_{WH}$ [-]</td>
<td>0.01</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.18</td>
<td>0.2</td>
</tr>
<tr>
<td>Soil</td>
<td>$FC$ [mm]</td>
<td>5</td>
<td>650</td>
<td>232.9</td>
<td>531.1</td>
<td>258.4</td>
<td>151</td>
<td>87.9</td>
</tr>
<tr>
<td></td>
<td>$L_p$ [-]</td>
<td>1</td>
<td>7</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.6</td>
</tr>
<tr>
<td></td>
<td>$\beta$ [-]</td>
<td>1</td>
<td>15</td>
<td>5.7</td>
<td>1.8</td>
<td>3</td>
<td>3.4</td>
<td>6.3</td>
</tr>
<tr>
<td></td>
<td>$PERC$ [mm/day]</td>
<td>0</td>
<td>10</td>
<td>0.72</td>
<td>5.34</td>
<td>1.4</td>
<td>0.86</td>
<td>3.12</td>
</tr>
<tr>
<td>Response and routing</td>
<td>$UZL$ [mm]</td>
<td>0</td>
<td>200</td>
<td>7.4</td>
<td>128.5</td>
<td>9.1</td>
<td>5.1</td>
<td>108.3</td>
</tr>
<tr>
<td></td>
<td>$K_{O}$ [1/day]</td>
<td>0.25</td>
<td>0.99</td>
<td>0.88</td>
<td>0.89</td>
<td>0.74</td>
<td>0.64</td>
<td>0.34</td>
</tr>
<tr>
<td></td>
<td>$K_{I}$ [1/day]</td>
<td>0.2</td>
<td>0.99</td>
<td>0.2</td>
<td>0.67</td>
<td>0.2</td>
<td>0.25</td>
<td>0.2</td>
</tr>
<tr>
<td></td>
<td>$K_{L}$ [1/day]</td>
<td>5·10^{-5}</td>
<td>0.1</td>
<td>0.003</td>
<td>0.017</td>
<td>0.037</td>
<td>0.02</td>
<td>0.013</td>
</tr>
<tr>
<td></td>
<td>MAXBAS [day]</td>
<td>1</td>
<td>10</td>
<td>2.7</td>
<td>5.1</td>
<td>2.7</td>
<td>2.7</td>
<td>2.1</td>
</tr>
<tr>
<td>Meteo</td>
<td>$P_{CALT}$ [%/100 m]</td>
<td>0</td>
<td>10</td>
<td>0.2</td>
<td>2.71</td>
<td>4.9</td>
<td>0.6</td>
<td>0.34</td>
</tr>
<tr>
<td></td>
<td>$T_{CALT}$ [C/100 m]</td>
<td>0.55</td>
<td>0.7</td>
<td>0.62</td>
<td>0.7</td>
<td>0.67</td>
<td>0.57</td>
<td>0.62</td>
</tr>
</tbody>
</table>
(1954–2010) and in 5 consecutive wettest, driest, warmest and coldest years, selected according to total annual precipitation depths and mean annual temperature observed at Valjevo meteorological station (Figure 3). In this way, five models are developed. For simplicity, these models are referred to as WET, DRY, WARM, COLD and RECORD.

The catchment is represented by a single parameter set, but it is delineated in three elevation zones of approximately same spans. Precipitation and temperature data are corrected with respect to the differences between mean zone elevation and the elevation of the Valjevo meteorological station.

The calibration period length is set to 5 years to enable identification of all parameters (Wagener et al. 2003) and a stable model calibration (Merz et al. 2011; Seibert & Vis 2012), except for the RECORD model, which uses the complete record. Each simulation started with the beginning of the water year. The first year of every simulation is intended for the model warm-up, what is considered sufficient to diminish the impact of the standard initial conditions in HBV-light (Seibert & Vis 2010). The computational time step in all simulations is 1 day.

The models are calibrated using the GAP optimisation algorithm. Initial parameter ranges (Table 1) are set after recommendations in the literature (e.g. Merz et al. 2011; Seibert & Vis 2012) and previous modelling experience of runoff modelling at this catchment (Todorović & Plavšić 2014). In this paper, the aggregate objective function (OF) is used for model calibration. OF consists of the Nash-Sutcliffe efficiency coefficients for flows (NSE, Nash & Sutcliffe 1970) and logarithms of flows (NSElog), and of volumetric error (VE). The NSE and VE are calculated as follows (Krause et al. 2005):

\[
\text{NSE} = 1 - \frac{\sum_i (Q_{\text{OBS},i} - Q_{\text{SIM},i})^2}{\sum_i (Q_{\text{OBS},i} - Q_{\text{OBS}})^2}
\]
\[
\text{VE} = 1 - \frac{\sum_i |Q_{\text{OBS},i} - Q_{\text{SIM},i}|}{\sum_i Q_{\text{OBS},i}}
\]

where \(Q_{\text{OBS},i}\) and \(Q_{\text{SIM},i}\) are the observed and simulated streamflows in time step \(i\), respectively.

Figure 3 | Annual precipitation and mean annual temperatures observed at the Valjevo meteorological station, along with their 5-year moving averages (MAS).
The three measures participate in OF with weights of 0.35, 0.35 and 0.3, respectively. Lumping several objective functions into a composite one introduces different aspects of model performance in the calibration procedure (Yilmaz et al. 2010). For example, NSE value primarily depends on model efficiency in high-flow domain (Krause et al. 2005; Oudin et al. 2006). Including log-transformed flows into OF enables quantifying model performance in the low-flow domain (Oudin et al. 2006). VE quantifies bias in the simulated flow volume, which is quite important in flow projections under CC (Merz et al. 2011; Coron et al. 2012). All the measures may take values from $-\infty$ (total lack of fit) to 1 (perfect fit), so their scaling is not necessary. Negative values of the performance measures indicate that mean observed flows would be better predictor than the model (Krause et al. 2005).

Every model is evaluated over the remaining four periods. Because some of the simulation periods overlap (wettest and coldest, the full record period with four sub-periods), this evaluation procedure should not be considered a genuine DSST (Coron et al. 2014).

Climate change impact on hydrologic regime

In this paper, the outputs from five GCM–RCM chains, presented in Langsholt et al. (2013), are used. These include ECHAM5R3-RACMO, ECHAM5R3-REMO, ECHAM5R3-RegCM3, HADCM3Q0-CLM and HADCM3Q0-HADRM3 Q0, all obtained under the A1B IPCC scenario. This scenario implies balanced reliance on fossil and non-fossil energy sources (IPCC 2000). The GCM outputs are down-scaled using RCMs with horizontal resolution of 50 or 25 km, depending on an individual RCM. The RCM outputs are bias-corrected for location of the Valjevo meteorological station over the baseline period (1961–1990) using the quantile mapping approach. In this approach, the bias correction functions are developed for each month from the distribution functions of observed and simulated daily precipitation depths and temperatures. Monthly precipitation and mean monthly temperatures resulting from the HADCM3Q0-HADRM3Q0 GCM–RCM chain are presented in Figure 4 for the baseline period, the near (2011–2040) and distant future (2041–2070). This particular GCM–RCM chain results in the largest increase in temperature, compared to other considered chains.

As there are no ET measurements, ET rates simulated by the GCM–RCM chains cannot be bias corrected, so they are considered unreliable. Since only temperature and precipitation projections were available, average monthly PET rates for two time slices were estimated using the temperature based Hamon equation. Lu et al. (2005) showed that the PET rates estimated using the Hamon method are comparable to the results of more complex methods (such as the Priestley-Taylor method). Oudin et al. (2005) demonstrated satisfactory performance of hydrologic models using the Hamon method for PET assessment. The Hamon equation for mean monthly PET rates reads (Lu et al. 2005):

$$\text{PET}_i = k \cdot 0.1651 \cdot L_d \cdot \rho_{SAT}$$

where $L_d$ is daytime length (in multiples of 12 hours), $\rho_{SAT}$ is saturated vapour density at the mean monthly temperature and $k$ is the calibration coefficient. The values of $k$
are inferred from the ratios between monthly PET rates calculated after Eagleman and Hamon for 1949–2010.

RESULTS AND DISCUSSION

The values of the optimised parameters for five different calibration periods are presented in Table 1. Obviously, the parameter values vary markedly with the calibration period. Some parameters of the snow routine (e.g. CFR, the refreezing coefficient) and the soil routine (e.g. FC, maximum soil water storage, and UZL, the threshold parameter) show high sensitivity to the calibration period. On the other hand, the estimates of some parameters (e.g. CWH, the snow water equivalent retained by the snow pack at melting, and Lp, the coefficient defining the reduction in ET rates) are shown to be consistent.

Model performance in different periods is presented in Table 2 (NSE coefficients for daily, log-transformed daily and mean monthly flows), and in Table 3 (volumetric error and composite objective function of daily flows, and annual flow volume bias). Underlined values in the tables are those from the calibration period for each model. It should be emphasised that this catchment is rather heterogeneous and that high model efficiency is rather difficult to obtain with a daily computational time step (e.g. Langsholt et al., 2013; Todorović & Plavšić, 2014).

The best performance is obtained for model calibration in the wettest period, probably due to the high information content of the observations. Similar performance level is achieved in the coldest period, which is to be expected considering the overlap with the wettest one. According to Wilby (2005) and Moriasi et al. (2007), NSE values obtained for model calibration in these two periods are considered good, and results in the warmest period are satisfactory. The poorest performance is noticed for the calibration in 1988–1992, implying the model’s limited ability to simulate properly the catchment’s behaviour during dry periods. This behaviour may be due to inadequate spatial resolution of the data and/or low parameter identifiability (Wagener et al. 2003). Model performance in the full record simulation is also unsatisfactory, likely due to the variety of hydrologic responses observed, which cannot be properly represented by a single parameter set. Presumably, low performance over dry sub-periods averages out the performance during the remaining of the full record period, resulting in overall unsatisfactory efficiency. This is consistent with the findings of Vaze et al. (2010) and Luo et al. (2012), who showed that model efficiency decreases with the length of calibration period. NSE values increase with the calculation time step (e.g. weekly, monthly) what corroborates previous research.

Table 2 | Nash-Sutcliffe coefficients for daily flows (NSE), log-transformed daily flows (NSE<sub>log</sub>) and monthly flows (NSE<sub>mn</sub>)

<table>
<thead>
<tr>
<th>Simulation period</th>
<th>Measure</th>
<th>WET</th>
<th>DRY</th>
<th>WARM</th>
<th>COLD</th>
<th>REC.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1977–1981 (wet)</td>
<td>NSE</td>
<td>0.67</td>
<td>0.19</td>
<td>0.49</td>
<td>0.64</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td>NSE&lt;sub&gt;log&lt;/sub&gt;</td>
<td>0.64</td>
<td>-0.06</td>
<td>0.47</td>
<td>0.65</td>
<td>0.54</td>
</tr>
<tr>
<td></td>
<td>NSE&lt;sub&gt;mn&lt;/sub&gt;</td>
<td>0.79</td>
<td>0.1</td>
<td>0.49</td>
<td>0.81</td>
<td>0.72</td>
</tr>
<tr>
<td>1988–1992 (dry)</td>
<td>NSE</td>
<td>-0.19</td>
<td>0.17</td>
<td>-0.18</td>
<td>-0.4</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>NSE&lt;sub&gt;log&lt;/sub&gt;</td>
<td>-0.2</td>
<td>0.28</td>
<td>-0.3</td>
<td>-0.14</td>
<td>-0.04</td>
</tr>
<tr>
<td></td>
<td>NSE&lt;sub&gt;mn&lt;/sub&gt;</td>
<td>-0.37</td>
<td>0.33</td>
<td>-0.51</td>
<td>-0.58</td>
<td>-0.12</td>
</tr>
<tr>
<td>2006–2010 (warm)</td>
<td>NSE</td>
<td>0.35</td>
<td>0.29</td>
<td>0.59</td>
<td>0.35</td>
<td>0.42</td>
</tr>
<tr>
<td></td>
<td>NSE&lt;sub&gt;log&lt;/sub&gt;</td>
<td>0.4</td>
<td>0.07</td>
<td>0.68</td>
<td>0.43</td>
<td>0.4</td>
</tr>
<tr>
<td></td>
<td>NSE&lt;sub&gt;mn&lt;/sub&gt;</td>
<td>0.63</td>
<td>0.39</td>
<td>0.8</td>
<td>0.68</td>
<td>0.64</td>
</tr>
<tr>
<td>1976–1980 (cold)</td>
<td>NSE</td>
<td>0.61</td>
<td>0.15</td>
<td>0.43</td>
<td>0.61</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>NSE&lt;sub&gt;log&lt;/sub&gt;</td>
<td>0.65</td>
<td>-0.09</td>
<td>0.4</td>
<td>0.67</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td>NSE&lt;sub&gt;mn&lt;/sub&gt;</td>
<td>0.81</td>
<td>0.09</td>
<td>0.4</td>
<td>0.79</td>
<td>0.69</td>
</tr>
<tr>
<td>1954–2010 (full record)</td>
<td>NSE</td>
<td>0</td>
<td>0.08</td>
<td>-0.04</td>
<td>-0.02</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>NSE&lt;sub&gt;log&lt;/sub&gt;</td>
<td>0.36</td>
<td>0.07</td>
<td>0.25</td>
<td>0.37</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>NSE&lt;sub&gt;mn&lt;/sub&gt;</td>
<td>0.32</td>
<td>0.21</td>
<td>0.13</td>
<td>0.28</td>
<td>0.36</td>
</tr>
</tbody>
</table>

Table 3 | Volumetric error (VE), composite objective function (OF) and relative annual runoff volume error (RVE) in per cent

<table>
<thead>
<tr>
<th>Simulation period</th>
<th>Measure</th>
<th>WET</th>
<th>DRY</th>
<th>WARM</th>
<th>COLD</th>
<th>REC.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1977–1981 (wet)</td>
<td>VE</td>
<td>0.61</td>
<td>0.29</td>
<td>0.48</td>
<td>0.63</td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td>OF</td>
<td>0.64</td>
<td>0.13</td>
<td>0.48</td>
<td>0.64</td>
<td>0.53</td>
</tr>
<tr>
<td></td>
<td>RVE</td>
<td>0</td>
<td>5</td>
<td>12</td>
<td>2</td>
<td>-4</td>
</tr>
<tr>
<td>1988–1992 (dry)</td>
<td>VE</td>
<td>0.28</td>
<td>0.47</td>
<td>0.29</td>
<td>0.27</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td>OF</td>
<td>-0.06</td>
<td>0.3</td>
<td>-0.08</td>
<td>-0.11</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>RVE</td>
<td>-24</td>
<td>0</td>
<td>2</td>
<td>-5</td>
<td>-5</td>
</tr>
<tr>
<td>2006–2010 (warm)</td>
<td>VE</td>
<td>0.42</td>
<td>0.5</td>
<td>0.55</td>
<td>0.42</td>
<td>0.42</td>
</tr>
<tr>
<td></td>
<td>OF</td>
<td>0.39</td>
<td>0.22</td>
<td>0.61</td>
<td>0.4</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>RVE</td>
<td>-7</td>
<td>0</td>
<td>5</td>
<td>-1</td>
<td>-2</td>
</tr>
<tr>
<td>1976–1980 (cold)</td>
<td>VE</td>
<td>0.61</td>
<td>0.3</td>
<td>0.46</td>
<td>0.63</td>
<td>0.54</td>
</tr>
<tr>
<td></td>
<td>OF</td>
<td>0.63</td>
<td>0.11</td>
<td>0.43</td>
<td>0.64</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td>RVE</td>
<td>-2</td>
<td>5</td>
<td>14</td>
<td>0</td>
<td>-1</td>
</tr>
<tr>
<td>1954–2010 (full record)</td>
<td>VE</td>
<td>0.39</td>
<td>0.27</td>
<td>0.35</td>
<td>0.37</td>
<td>0.4</td>
</tr>
<tr>
<td></td>
<td>OF</td>
<td>0.24</td>
<td>0.14</td>
<td>0.17</td>
<td>0.24</td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td>RVE</td>
<td>-7</td>
<td>2</td>
<td>9</td>
<td>2</td>
<td>0</td>
</tr>
</tbody>
</table>
(e.g. Moriasi et al. 2007). For example, NSE calculated with mean monthly flows is almost twice the NSE with daily flows in the driest period. As shown by relative volumetric error (RVE) in Table 3, all models replicate the observed runoff volumes in the calibration period.

Generally, model performance in the evaluation periods decreases (Tables 2 and 3). The most pronounced loss of efficiency is exhibited in model application over the driest and full record periods. The WET model application in the driest period caused a substantial bias in the simulated runoff volume and markedly greater decrease in NSE values than the DRY model evaluation over the wettest period. This is consistent to the previous research (e.g. Vaze et al. 2010; Li et al. 2012), which demonstrated larger efficiency decrease if the calibration period were wetter than the evaluation one, than other way round. Interestingly, the DRY and RECORD models performed better in some evaluation periods than in the calibration ones.

The WET model efficiency in the coldest period is as good as the efficiency of the COLD model in the wettest one (Table 2). As parameters of these two models differ (Table 1), similar evaluation skill proves the equifinality assumption.

A modest decrease when exchanging parameters between the warmest and the coldest period indicates that temperature is not as limiting for model transferability as precipitation (note that the precipitation amount in these periods is similar). In other words, hydrologic simulations in a period contrasted to the calibration one in terms of temperature can be considered reliable, given similarity in precipitation.

Table 4 | Characteristic simulated hydrologic variables for the baseline period (1961–1990) and for two future time slices (Qp% is the flow exceeded for p% of time during a year)

<table>
<thead>
<tr>
<th>Time slice</th>
<th>Model</th>
<th>WET</th>
<th>DRY</th>
<th>WARM</th>
<th>COLD</th>
<th>RECORD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1961–1990</td>
<td>Precipitation [mm/year]</td>
<td>788.7</td>
<td>322.3</td>
<td>344.1</td>
<td>320.2</td>
<td>308.8</td>
</tr>
<tr>
<td></td>
<td>Runoff [mm/year]</td>
<td>293.7</td>
<td>20.1</td>
<td>32.2</td>
<td>32.7</td>
<td>24.0</td>
</tr>
<tr>
<td></td>
<td>Q5% [m³/s]</td>
<td>29.3</td>
<td>7.6</td>
<td>5.9</td>
<td>5.1</td>
<td>6.7</td>
</tr>
<tr>
<td></td>
<td>Q50% [m³/s]</td>
<td>4.0</td>
<td>7.6</td>
<td>5.9</td>
<td>5.1</td>
<td>6.7</td>
</tr>
<tr>
<td></td>
<td>Q95% [m³/s]</td>
<td>2.2</td>
<td>2.7</td>
<td>0.7</td>
<td>1.4</td>
<td>1.68</td>
</tr>
<tr>
<td></td>
<td>ET [mm/year]</td>
<td>524.9</td>
<td>555.5</td>
<td>541.2</td>
<td>495.9</td>
<td>473.9</td>
</tr>
<tr>
<td>2011–2040</td>
<td>Projected precipitation [mm/year]</td>
<td>846.4</td>
<td>356.3</td>
<td>350.9</td>
<td>351.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Runoff [mm/year]</td>
<td>309.4</td>
<td>21.0</td>
<td>37.0</td>
<td>37.8</td>
<td>31.9</td>
</tr>
<tr>
<td></td>
<td>Q5% [m³/s]</td>
<td>32.4</td>
<td>6.7</td>
<td>5.8</td>
<td>4.8</td>
<td>6.9</td>
</tr>
<tr>
<td></td>
<td>Q50% [m³/s]</td>
<td>3.8</td>
<td>2.1</td>
<td>0.6</td>
<td>1.0</td>
<td>1.7</td>
</tr>
<tr>
<td></td>
<td>Q95% [m³/s]</td>
<td>2.0</td>
<td>1.5</td>
<td>0.2</td>
<td>1.3</td>
<td>1.3</td>
</tr>
<tr>
<td></td>
<td>Evapotranspiration [mm/year]</td>
<td>556.2</td>
<td>586.9</td>
<td>513.1</td>
<td>488.7</td>
<td></td>
</tr>
<tr>
<td>2041–2070</td>
<td>Projected precipitation [mm/year]</td>
<td>793.4</td>
<td>307.8</td>
<td>312.2</td>
<td>322.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Runoff [mm/year]</td>
<td>259.9</td>
<td>18.0</td>
<td>30.7</td>
<td>33.2</td>
<td>29.0</td>
</tr>
<tr>
<td></td>
<td>Q5% [m³/s]</td>
<td>26.4</td>
<td>5.4</td>
<td>4.8</td>
<td>4.3</td>
<td>6.3</td>
</tr>
<tr>
<td></td>
<td>Q50% [m³/s]</td>
<td>3.2</td>
<td>1.5</td>
<td>0.2</td>
<td>1.3</td>
<td>1.3</td>
</tr>
<tr>
<td></td>
<td>ET [mm/year]</td>
<td>549</td>
<td>580.2</td>
<td>497.9</td>
<td>467.6</td>
<td></td>
</tr>
</tbody>
</table>

Total annual runoff, high, median and low flows (representing 5, 50 and 95 exceedance percentiles, respectively), and ET for the considered future periods (2011–2040 and 2041–2070), calculated using the outputs of HADCM3Q0-HADRM3Q0 GCM–RCM as the input for calibrated models, are presented in Table 4. Changes in the simulated variables over the future time slices, relative to variables simulated over the baseline period (1961–1990), are illustrated in Figure 5(a) and 5(b).

The results indicate that an increase in mean (total) annual runoff and high flows can be expected in the near future (2011–2040). The magnitude of changes varies with the hydrologic model applied. For example, increase in high flows varies between 10.4 (WET) and 33.1% (RECORD). As for the remaining variables, projected changes vary in sign and magnitude with the model used. Similar behaviour is observed over the second time slice (2041–2070), although variability among flow projections is somewhat larger compared to the first one. Similar
patterns in the relative changes are noticed if the outputs of other GCM–RCM chains are considered (not shown here).

Relative changes in the hydrologic projections in the near future, obtained using the model with the best calibration performance (WET) being forced with the outputs of five different GCM–RCMs, are presented in Figure 5(c). The variability in the projections due to selection of the calibration period is comparable in magnitude to the variability due to the GCM–RCM chain. For example, changes in low flows in the near future vary between 2.7 (RECORD) and −24.4% (COLD), and between 7.7 (ECHAM5R3-RegCM3) and −26.8% (ECHAM5R3-RACMO). Similar patterns are obtained with other hydrologic models (not shown here). These results show that selection of period for model calibration introduces uncertainty to the hydrologic projections under CC.

Low efficiency models (e.g. the DRY model) usually would not be applied in CC impact analysis. However, the aim of this study is not to develop flow projections, but to investigate how the calibration period contributes to uncertainty in hydrologic projections under CC.

CONCLUSIONS

The conceptual HBV-light hydrologic model of the Kolubara River catchment, aimed at continuous runoff simulations, is calibrated using a composite objective function and combined global and local optimisation algorithms. The model is calibrated against daily flows observed at the Slovac stream gauge in the full hydrologic record period (1954–2010), and over 5 consecutive wettest,
driest, warmest and coldest years observed at the Valjevo meteorological station. In this way, five models are obtained (WET, DRY, WARM, COLD and RECORD).

The model parameters are shown to be heavily dependent on the calibration period. Some parameters of the snow and soil routines are the most sensitive towards climate properties of the calibration period. Similar performance of the WET and COLD models on the coldest and the wettest periods, respectively, proves the equifinality of different parameter sets.

The best model performance is obtained in the wettest and coldest calibration periods, while the poorest is noticed over the driest and the full record periods. The DRY and RECORD models perform better in some evaluation periods than in their calibration periods. This indicates the model inability to replicate the catchment’s behaviour over dry periods for the given spatial and temporal data resolution. The highest drop in model efficiency is noticed for model calibration over the wettest period and evaluation in the driest one. Parameter transfer between the warmest and coldest periods caused a moderate but acceptable decrease in model performance, suggesting that models calibrated over colder periods may reproduce a catchment’s behaviour over the warmer ones, especially in terms of runoff volume. This is rather significant when it comes to the application of hydrologic models for assessment of CC effects. These results also imply that large uncertainties in simulated flows can be expected if the model is applied over periods drier than the calibration one.

Five calibrated models are applied to develop hydrologic projections in the near future (2011–2040) and distant future (2041–2070), using the outputs of five GCM–RCM chains under A1B emission scenario. Projected variables are compared to the corresponding ones simulated over the baseline period (1961–1990). Changes in the simulated variables vary to a considerable extent with the hydrologic model used, especially in the distant future. The degree of variability of the projected changes due to different calibration period is comparable in magnitude to the variability due to different GCM–RCM chains.

This paper demonstrates that the choice of a period for hydrologic model calibration can significantly contribute to the uncertainty of hydrologic projections under CC. Model robustness, i.e. efficiency over various climatically contrasted periods, should be evaluated prior to model application for the development of hydrologic projections. The results reveal that departures from temperatures in the calibration period are not expected to cause unreliable flow projections, as opposed to precipitation depths. Models calibrated over periods similar to the projected conditions in terms of precipitation amounts may result in less uncertain projections. Simulations over prolonged dry periods (as is projected for spring and summer for the Kolubara catchment in the distant future), are expected to be rather uncertain. These conclusions are to be further confirmed in future research by involving additional sources of uncertainties related to climate models, such as different GHG emission scenarios or different methods for downscaling and bias correction of the climate model outputs.

Inconsistency in hydrologic model performance and model parameter transferability among different (especially climatically contrasted) periods have been investigated for decades. However, further research is needed to achieve a more consistent hydrologic model performance. This should be done by, e.g. employing model ensembles, which are shown to be more transferable compared to the individual hydrologic models (Seiller et al. 2022), or model weighting methods (e.g. Oudin et al. 2006). Consistency in model performance would mitigate the uncertainties and contribute to more reliable hydrologic projections.

ACKNOWLEDGEMENTS

The research presented is supported by the Ministry of Science and Technological Development of the Republic of Serbia (project TR37005). Data used are provided by Republic Hydrometeorological Service of Serbia.

REFERENCES


Osuch, M., Romanowicz, R. J. & Booij, M. J. 2014 The influence of parametric uncertainty on the relationships between HBV model parameters and climatic characteristics. Hydrological Sciences Journal.


Xu, C.-Y. 2000 Climate change and hydrologic models: a review of existing gaps and recent research developments. Water Resources Management 13, 509–582.


First received 25 June 2014; accepted in revised form 29 April 2015. Available online 15 June 2015.