Hydrological Sciences Journal

Stability of model performance and parameter values on two catchments facing changes in climatic conditions

Hong Li, Stein Beldring & Chong-Yu Xu

To cite this article: Hong Li, Stein Beldring & Chong-Yu Xu (2015) Stability of model performance and parameter values on two catchments facing changes in climatic conditions, Hydrological Sciences Journal, 60:7-8, 1317-1330, DOI: 10.1080/02626667.2014.978333

To link to this article: http://dx.doi.org/10.1080/02626667.2014.978333

Accepted online: 23 Oct 2014. Published online: 12 Aug 2015.

Submit your article to this journal

Article views: 101

View related articles

View Crossmark data

Citing articles: 1 View citing articles
Stability of model performance and parameter values on two catchments facing changes in climatic conditions

Hong Li\textsuperscript{1}, Stein Beldring\textsuperscript{2} and Chong-Yu Xu\textsuperscript{1}

\textsuperscript{1}Department of Geosciences, University of Oslo, Blindern, N-0316 Oslo, Norway
chongyu.xu@geo.uio.no
\textsuperscript{2}Norwegian Water Resources and Energy Directorate, Majorstua, N-0301 Oslo, Norway

Received 5 April 2014; accepted 11 September 2014

Editor D. Koutsoyiannis; Guest editor G. Thirel

Abstract Hydrological models are often used for studying the hydrological effects of climate change; however, the stability of model performance and parameter values under changing climate conditions has seldom been evaluated and compared. In this study, three widely-used rainfall–runoff models, namely the SimHYD model, the HBV model and the Xin'anjiang model, are evaluated on two catchments subject to changing climate conditions. Evaluation is carried out with respect to the stability in their performance and parameter values in different calibration periods. The results show that (a) stability of model performance and parameter values depends on model structure as well as the climate of catchments, and the models with higher performance scores are more stable in changing conditions; (b) all the tested models perform better on a humid catchment than on an arid catchment; (c) parameter values are also more stable on a humid catchment than on an arid catchment; and (d) the differences in stability among models are somewhat larger in terms of model efficiency than in model parameter values.

Key words climate change; conceptual models; hydrological regime; non-stationarity; rainfall–runoff models

1 INTRODUCTION

Climate change is one of the most serious environmental threats that humanity has ever been confronted with (Hofgaard 1997, Xu 1999a, Xu et al. 2005, O’Brien et al. 2006). To comprehend the hydrological impacts of climate change, scientists use hydrological models driven by different climate change scenarios. Hydrological models transform climate conditions to hydrological responses at catchment, regional and global scales (Widén-Nilsson et al. 2007, Kizza et al. 2013, Li et al. 2013). Therefore, projections can be made for water resources and floods under future climate conditions by hydrological models taking inputs from global or regional climate models.

Rainfall–runoff models are mathematical tools that describe the relationship between the precipitation
over a catchment area or a region and the resulting runoff by means of mathematical equations. The constants of these equations, i.e., parameters of the models, are estimated based on observations of precipitation and runoff with the help of other meteorological variables and basin physical data. However, climate and land-use change can change or modify the rainfall—runoff relationship (Lavabre et al. 1993, Zhang et al. 2011b). The most distinct indicator of climate change is the increase in global average temperature and/or the change in precipitation. Runoff is simultaneously modified by climate change (e.g., precipitation and temperature) and human activities (land use, water consumption and construction of dams) (Milly et al. 2008, Zhang et al. 2011a, Zeng et al. 2014).

In mathematics, strict stationarity of a time series requires that all moments of its probability distribution function are identical across time, whereas a time series is non-stationary if one of its moments is not constant over time. A less strict type of stationarity, called ‘weak’ or ‘second-order’ stationarity, is that in which the first- and second-order moments are constant over time (Chen and Rao 2002, Murphy and Ellis 2013). Due to the limited length of hydrological records, normally 50 years or less, only the first and second order moments can be reliably estimated. Therefore, a stationary hydrological time series is actually only weakly stationary in mathematical terms (Chen and Rao 2002).

To substantiate applicability of hydrological models in changing conditions, the models have to pass rigorous tests. Klemes (1986) prescribed a differential split-sample method for testing the transferability of hydrological models under non-stationary conditions. Time series of data are subdivided into two sub-periods with contrasting climatic conditions, like warm and cold, dry and wet. Validation on the period that has different climate conditions from the calibration period can test the dependence of model parameters on climate and transferability of parameters in various climate conditions (Brigode et al. 2013). Using this method, Xu (1999b) evaluated the WASMOD model on two Swedish catchments with two sub-periods having different climate conditions. Seibert (2003) used this method to assess the ability of the HBV model on simulating peak flow on four Swedish catchments. Li et al. (2012) tested the DWBM model and the SimHYD model on 30 unpaired catchments in Australia with two sub-periods of changing climate conditions. In these studies, only two sub-periods classified as dry or wet act as calibration and verification by turn and these studies show deterioration in model performance of validations.

If longer records of data are available, this test can be generalized or modified to multiple calibration periods and yields multiple optimized parameter sets. The statistics of the optimized parameter values and all possible validations in independent periods give insights about uncertainty in calibrations and stability of parameters (Coron et al. 2012, Brigade et al. 2013). This test can be seen as an extension or modification of the differential split-sample test (Coron et al. 2012). Thus the modified differential split-sample test method provides a possibility for examining model stability under changing climate conditions.

In general, parameter values in dynamic models are assumed to represent stable catchment conditions, while rainfall, temperature and other inputs are time-varying boundary conditions (Merz et al. 2011). Therefore, calibrated parameters are expected to be insensitive to changes in climatic conditions. This is a fundamental assumption for any hydrological model to be used in simulating climate change impact on hydrological variables (Schulze 2000, Zeng et al. 2014). However, using the aforementioned method, Merz et al. (2011) found that parameters of snow and soil moisture are more sensitive to climatic conditions of calibration periods than other parameters. Moreover, Coron et al. (2012) found that the transferability of model parameters is more affected by a change in precipitation than in temperature and potential evaporation by three models of different structures. However, no universal conclusions can be drawn and further studies are always recommended.

A large number of rainfall—runoff models of varying degrees of complexity have been developed, ranging from empirical black-box models to conceptual quasi-physical models and physically-based distributed models (Shamseldin et al. 1997, Singh 2002, Xu and Singh 2004). Comparisons of models with different degrees of complexity demonstrated that there is no single model that performs better than other models for all types of catchments and under all circumstances (Refsøgaard and Knudsen 1996, Shamseldin et al. 1997, Butts et al. 2004, Georgakakos et al. 2004, Reed et al. 2004, Duan et al. 2007, Habets et al. 2009). Furthermore, hydrological simulations of different water balance components by different models are different even though runoff simulations exhibit the same pattern. Alley (1984) was probably the first to notice the ‘invisible’ discrepancies. He found that after calibration,
different monthly models reproduced similar runoff series, but simulated values of state variables such as soil moisture storage were substantially different among models. More recently, using six conceptual models, Jiang et al. (2007) demonstrated that even though these six models had similar capabilities in reproducing historical water balance components, greater differences occurred in simulating impacts of the postulated climate changes on different water balance components.

The above mentioned studies revealed that each model has its own strengths in capturing different aspects of real world processes (Duan et al. 2007). No model is likely to perform satisfactorily at all times or under all conditions (e.g. perhaps not all of its structural assumptions are valid or the conditions under which it is assumed to operate are not entirely fulfilled) (Shamseldin et al. 1997). It is, therefore, crucially important to investigate the different performance of the models under different conditions and to explore the reasons for the differences among the models.

To involve as many models as possible in studies of non-stationary hydrological relationships, a workshop entitled “Testing simulation and forecasting models in non-stationary conditions” was organized as a special session during the IAHS/IAPSO/IASPEI Joint Assembly in Gothenburg, Sweden in July 2013. Time series of precipitation, temperature, discharge and other variables of 14 catchments with significant changes in land cover, climate and dam construction were provided as a common database. Twenty four models, including black-box models, conceptual models and physically-based models, were set-up on these catchments by research groups around the world. The modellers were asked to follow a similar calibration and validation protocol and results were analysed using common criteria, such as the Nash-Sutcliffe efficiency and the bias (Thirel et al. 2015). Besides these common criteria, participants also put forward their own insights and solutions to hydrological modelling on the catchments.

This paper originates from one of the 30 presentations made at the workshop, and follows the guidance of the workshop. As a contribution to the session, the objective of this study is to compare and quantify the differences in model performance and stabilities of parameters of three widely-used rainfall–runoff models in estimating runoff and water balance for two catchments located in very different climate zones. In the two catchments, distinct changes are detected in precipitation and runoff.

2 STUDIED CATCHMENTS

2.1 The Lianshui catchment

The Lianshui catchment at the Xiangxiang gauging station is located in the humid zone in southern China (Fig. 1). The mean altitude is 209 m a.m.s.l. ranging from 52 to 1033 m a.m.s.l. The upstream part of the catchment is mountainous (altitude >200 m) and comprises approximately 36% of the total area. The midstream area is dominated by hills (100 m < altitude ≤ 200 m), and covers around 55% of the catchment. Plains in the downstream (altitude ≤ 100 m) cover the rest of the area. As for land-use, forest and cropland, cover 48 and 47% of the total area of the catchment, respectively. This catchment is heavily influenced by a monsoon climate, which brings heavy rainfall from the south. Annual precipitation is around 1360 mm year\(^{-1}\); yearly average temperature is 17°C; and the annual runoff is 660 mm. The rainy season is from April to September accounting for almost 70% of annual precipitation.


2.2 The Wimmera catchment

The Wimmera catchment at the Glenorchy Weir Tail gauging station is located in the arid zone in southeast Australia (Fig. 2). The mean altitude is 300 m a.m.s.l. Around 13% of the total area is above 413 m, mainly located in the east; approximately 83% of the total area is between 187 and 413 m. Cropland covers 60% of the total area, with forests and grassland over the remainder (Hansen et al. 2000). Annual precipitation is 560 mm year\(^{-1}\); yearly average temperature is 13.5°C; and annual runoff is 40 mm.

Time series of precipitation, temperature, potential evaporation and runoff at a daily time step from 1960 to 2005 were used. The period from 1960 to 1965 was kept for warming up the models. There were five sub-periods of 8-years each: 1966–1973 (P1), 1974–1981 (P2), 1982–1989 (P3), 1990–1997 (P4), and 1998–2005 (P5). Six years before the start of each sub-period are used for model warming up.
The division of sub-periods is the same as in Thirel et al. (2015), except that the whole period is defined from 1966 to 2005. Very low runoff was recorded from 1997 to 2005. These years, called the “Millennium Drought”, completely covered sub-period P5. Because the Wimmera catchment is located in an arid region and as the climate has a larger inter-annual variability than the Lianshui catchment, a longer sub-period and warming-up period were used in model evaluation.

The area of the Lianshui catchment is twice as large as the Wimmera catchment; it also experienced twice as much as precipitation. However, the Lianshui catchment has a six times larger runoff coefficient than the Wimmera catchment as shown in Fig. 3. More information about the two catchments is tabulated in Table 1.

3 METHODS AND MODELS

In this study three models were compared on the two catchments. The models are programmed in different computer languages and calibrated by different algorithms. Successful applications of these models by others and ourselves have verified the merits of the models and their calibration packages. Since the main objective of the study is not to compare the calibration routines, differences induced by the calibration algorithms are not considered. The calibration routines are discussed in Sections 3.2–3.4.

Moreover, the results presented herein are obtained with different initial parameter and state values in order to ensure that globally optimal parameter values are obtained. Occasionally the optimization results are not satisfactory with limited predefined initial parameter values especially on the Wimmera catchment.

Fig. 1 Altitude and location of the Lianshui catchment at the Xiangxiang gauging station. The black line is the Lianshui River and the black dot is the Xiangxiang gauging station.

Fig. 2 Altitude and location of the Wimmera catchment at the Glenorchy Weir Tail gauging station. The black line is the Wimmera River and the black dot is the Glenorchy Weir Tail gauging station.

Fig. 3 Mean annual precipitation (grey bars) and runoff coefficients (solid line) during each sub-period in the Lianshui catchment (left) and the Wimmera catchment (right). The bold dashed lines are the average annual precipitation for the whole period, and the dotted lines are the average runoff coefficients for the whole period. The ticks on the x-axis are the starting years of the sub-periods.
3.1 Model performance criteria

The objective function of calibration used in the study is the Nash-Sutcliffe coefficient (NSE) (Nash and Sutcliffe 1970). Even though this criterion is sensitive to the peak and underestimates the variability (Gupta et al. 2009, Vansteenkiste et al. 2014), it is still the most widely used objective function in general hydrological modelling (Perrin et al. 2001, Chiew and Siriwardena 2005, Pechlivanidis et al. 2011, Pushpalatha et al. 2012). In the model evaluation, two criteria, i.e. NSE and the relative mean error (RME), are adopted as requested in the guidance of the workshop. Their formulas are shown in Table 2.

If N is the number of sub-periods, the models are first calibrated on each sub-period resulting in N sets of parameter values and model performance scores. Each calibration is then validated in all other sub-periods. Thus, there are N calibrations and N*(N-1) validations for the sub-periods plus one calibration for the whole period.

3.2 The SimHYD model

The SimHYD model is a lumped conceptual rainfall–runoff model. It simulates daily runoff (surface runoff and baseflow) using daily precipitation and potential evaporation as input data. This model is one of the most commonly used hydrological models in Australia for climate change studies (Jones et al. 2006). More details and algorithms of the model can be found in Chiew and Siriwardena (2005).

Nine parameters need to be calibrated and the five most sensitive parameters (Chiew and Siriwardena 2005) plus correction factors of rainfall and potential evaporation are selected to analyse their stabilities in different calibration periods. The ranges of the parameters are primarily determined from Chiew and Siriwardena (2005) and adjusted accordingly in the study to better suit the two catchments. The final parameter ranges are tabulated in Table 3.

The SimHYD in Fortran is automatically calibrated by a global optimization procedure entitled the shuffled complex evolution (SCE-UA) method developed by Duan et al. (1992). The SCE-UA method indeed estimates the global optimization values (Duan et al. 1992, Doherty and Johnston 2003, Coron et al. 2012).

Table 1 Summary information of the two catchments. No. refers to the number of sub-periods and Length is the duration of each sub-period.

<table>
<thead>
<tr>
<th>Catchment</th>
<th>(Longitude, Latitude)</th>
<th>Area (km²)</th>
<th>Whole period</th>
<th>No.</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lianshui</td>
<td>(112.52°E, 27.72°N)</td>
<td>5499</td>
<td>1964–2003</td>
<td>8</td>
<td>5 years</td>
</tr>
<tr>
<td>Wimmera</td>
<td>(142.79°E, 36.98°S)</td>
<td>2000</td>
<td>1966–2005</td>
<td>5</td>
<td>8 years</td>
</tr>
</tbody>
</table>

Table 2 Evaluation criteria and their corresponding formulations (Oi and Si are the observed and simulated flow, respectively; i is the time series index; and n is the total number of time steps).

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Formula</th>
<th>Range</th>
<th>Perfect value</th>
</tr>
</thead>
<tbody>
<tr>
<td>NSE</td>
<td>[ \frac{\sum_{i=1}^{n} (O_i - S_i)^2}{1 - \frac{\sum_{i=1}^{n} (O_i - \bar{O})^2}{\sum_{i=1}^{n} O_i}} ]</td>
<td>(-∞, 1]</td>
<td>1</td>
</tr>
<tr>
<td>RME</td>
<td>[ \frac{\sum_{i=1}^{n} (S_i - O_i)}{\sum_{i=1}^{n} O_i} ]</td>
<td>(-∞, +∞)</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 3 Meanings and ranges of nine parameters in the SimHYD model. The analysed parameters are underlined.

<table>
<thead>
<tr>
<th>Index</th>
<th>Parameter</th>
<th>Meaning (unit)</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>INSC</td>
<td>Interception storage capacity (mm)</td>
<td>[0, 20]</td>
</tr>
<tr>
<td>2</td>
<td>SMSC</td>
<td>Soil moisture storage capacity (mm)</td>
<td>[20, 500]</td>
</tr>
<tr>
<td>3</td>
<td>SUB</td>
<td>Constant of proportionality in interflow equation</td>
<td>[0, 1]</td>
</tr>
<tr>
<td>4</td>
<td>CRAK</td>
<td>Constant of proportionality in groundwater recharge equation</td>
<td>[0, 1]</td>
</tr>
<tr>
<td>5</td>
<td>K</td>
<td>Log_10 of baseflow linear recession parameter</td>
<td>[-2.5, -0.5]</td>
</tr>
<tr>
<td>6</td>
<td>SQ</td>
<td>Infiltration loss exponent</td>
<td>[0, 6]</td>
</tr>
<tr>
<td>7</td>
<td>COEFF</td>
<td>Maximum infiltration loss (mm)</td>
<td>[20, 400]</td>
</tr>
<tr>
<td>8</td>
<td>PC</td>
<td>Rainfall multiplier</td>
<td>[0.5, 2.0]</td>
</tr>
<tr>
<td>9</td>
<td>EC</td>
<td>Potential evaporation multiplier</td>
<td>[0.2, 5.0]</td>
</tr>
</tbody>
</table>
3.3 The HBV model

The HBV model is a conceptual rainfall–runoff model aimed at the Scandinavian countries (Bergström 1992, Lindström et al. 1997). The model is also widely used in other countries (more than 80 worldwide), especially in northern Europe. The model version used in this study was developed by Beldring et al. (2003). The inputs are daily precipitation and air temperature; the outputs are daily runoff, groundwater and other water balance components. More details of the model can be found in Li et al. (2014).

Thirteen parameters need to be calibrated (Table 4) and the seven most sensitive parameters (Seibert 1997, 1999) are selected to analyse their stabilities. The ranges of the parameters were primarily determined by Seibert (1997) and Li et al. (2014), and were adjusted accordingly in the study to better suit the two catchments. The final parameter ranges are tabulated in Table 4.

The HBV model in C++ is calibrated by a free package called PEST (Model-Independent Parameter Estimation & Uncertainty Analysis) (Doherty and Johnston 2003). It is based on the Gauss-Marquardt-Levenberg algorithm, which combines advantages of the inverse Hessian matrix and the steepest gradient method to allow a fast and efficient convergence towards to the best value of the objective function (Doherty and Johnston 2003, Wrede et al. 2013).

3.4 The Xin’anjiang model

The Xin’anjiang model (XAJ) was developed in 1973 (Zhao 1992). The model has been applied successfully over a very large range of areas including the agricultural, pastoral and forested lands of China except the loess (Zhang and Lindström 1996, Jiang et al. 2007, Xu et al. 2013). The inputs are daily precipitation and potential evaporation and the outputs are daily runoff and other water balance components.

Fifteen parameters need to be calibrated and seven of them are sensitive parameters as recommended by Zhao (1992). The ranges of the parameters are determined by the study of Chen et al. (2012) and are adopted herein. The parameters underlined in Table 5 are analysed for time stability.

The XAJ model in Fortran is optimized by three algorithms (Chen et al. 2012), namely the Rosenbrock, simplex and genetic algorithms (Wang 1991). Chen et al. (2012) successfully applied the XAJ model as well as the calibration routines to study the impacts of climate change on runoff on the Hanjiang basin in China.

4 RESULTS

4.1 Changes of model performance in different sub-periods

4.1.1 The Lianshui catchment  

The results of the RME values are presented in Fig. 4. It is seen that as far as RME is concerned, the XAJ model and the HBV model perform better than the SimHYD model on the Lianshui catchment, since the RME values concentrate around zero and no significant bias is seen. The SimHYD model shows negatively biased results.

Results of the NSE values are shown in Fig. 5. It shows that the XAJ model gives the highest NSE values both in calibrations and validations, and the SimHYD model gives the lowest NSE values.

Table 4 Meanings and ranges of 13 parameters in the HBV model. The analysed parameters are underlined.

<table>
<thead>
<tr>
<th>Index</th>
<th>Parameter</th>
<th>Meaning (unit)</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>PREC</td>
<td>Precipitation correction</td>
<td>[0.5, 2.0]</td>
</tr>
<tr>
<td>2</td>
<td>INTER_MAX</td>
<td>Maximum interception storage (m)</td>
<td>[0.001, 0.100]</td>
</tr>
<tr>
<td>3</td>
<td>EPO</td>
<td>Potential evaporation capacity (m/d)</td>
<td>[10^{-4}, 5 × 10^{-3}]</td>
</tr>
<tr>
<td>4</td>
<td>WETCORR</td>
<td>Transpiration correction</td>
<td>[0, 1]</td>
</tr>
<tr>
<td>5</td>
<td>FCA</td>
<td>Field capacity (m)</td>
<td>[0.05, 1.50]</td>
</tr>
<tr>
<td>6</td>
<td>FCD</td>
<td>Maximum evaporation efficiency</td>
<td>[0.005, 1.000]</td>
</tr>
<tr>
<td>7</td>
<td>BETA</td>
<td>Shape coefficient of soil moisture</td>
<td>[0.5, 15.0]</td>
</tr>
<tr>
<td>8</td>
<td>INFM</td>
<td>Infiltration capacity (m)</td>
<td>[0.5, 1.5]</td>
</tr>
<tr>
<td>9</td>
<td>KUZ</td>
<td>Recession coefficient of the upper zone</td>
<td>[0.05, 1.00]</td>
</tr>
<tr>
<td>10</td>
<td>ALFA</td>
<td>Nonlinear drainage coefficient of the upper zone</td>
<td>[0.2, 2.0]</td>
</tr>
<tr>
<td>11</td>
<td>PERC</td>
<td>Percollation from upper zone to lower zone</td>
<td>[0, 1]</td>
</tr>
<tr>
<td>12</td>
<td>KLZ</td>
<td>Recession coefficient of the lower zone</td>
<td>[0, 1]</td>
</tr>
<tr>
<td>13</td>
<td>DRAW</td>
<td>Draw up coefficient from lower zone to soil moisture</td>
<td>[0, 1]</td>
</tr>
</tbody>
</table>
The NSE range of validations (the length of the red bars in Fig. 5) is a measure of the sensitivity of models to climate conditions. There are seven validations for each sub-period. The ranges are the largest in the SimHYD model. The performance scores of the HBV model and the XAJ model are similar. Additionally, the length of the ranges reflects to some extent the problem of parameter equifinality and uncertainty. The range bars are very short, especially those of the HBV model and the XAJ model, indicating that the uncertainty caused by parameters is low.

Normally, the highest NSE is obtained during the calibration. However, in the XAJ model, in some sub-periods, for example, P2 and P6, the maximum NSE values of the validations are slightly larger than those of the calibration. This is mainly caused by the fact that the calibration procedures

Table 5 Meanings and ranges of 15 parameters in the XAJ model. The sensitive parameters are underlined.

<table>
<thead>
<tr>
<th>Index</th>
<th>Parameter</th>
<th>Meaning (unit)</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>WM</td>
<td>Areal tension water capacity (mm)</td>
<td>[50, 250]</td>
</tr>
<tr>
<td>2</td>
<td>X</td>
<td>Fraction of upper water in WM</td>
<td>[0, 1]</td>
</tr>
<tr>
<td>3</td>
<td>Y</td>
<td>Fraction of upper water in (1–X) × WM</td>
<td>[0.1, 1.5]</td>
</tr>
<tr>
<td>4</td>
<td>KE</td>
<td>Evaporation coefficient</td>
<td>[0.5, 1.5]</td>
</tr>
<tr>
<td>5</td>
<td>B</td>
<td>Tension water distribution index</td>
<td>[0.01, 1.00]</td>
</tr>
<tr>
<td>6</td>
<td>SM</td>
<td>Areal free water capacity (mm)</td>
<td>[5, 100]</td>
</tr>
<tr>
<td>7</td>
<td>EX</td>
<td>Free water distribution index</td>
<td>[0.05, 50.00]</td>
</tr>
<tr>
<td>8</td>
<td>CI</td>
<td>Fraction of free water to interflow</td>
<td>[0, 1]</td>
</tr>
<tr>
<td>9</td>
<td>CG</td>
<td>Fraction of free water to groundwater</td>
<td>[0, 1]</td>
</tr>
<tr>
<td>10</td>
<td>IMP</td>
<td>Impermeable coefficient</td>
<td>[0, 1]</td>
</tr>
<tr>
<td>11</td>
<td>C</td>
<td>Deep layer evaporation coefficient</td>
<td>[0, 1]</td>
</tr>
<tr>
<td>12</td>
<td>KI</td>
<td>Interflow recession coefficient</td>
<td>[0, 1]</td>
</tr>
<tr>
<td>13</td>
<td>KG</td>
<td>Groundwater recession coefficient</td>
<td>[0.5, 1]</td>
</tr>
<tr>
<td>14</td>
<td>N</td>
<td>Number of reservoirs of the Nash model</td>
<td>[1, 10]</td>
</tr>
<tr>
<td>15</td>
<td>NK</td>
<td>Storage constant of the Nash model</td>
<td>[0.5, 50]</td>
</tr>
</tbody>
</table>

Fig. 4 Histogram of the RME values on the Lianshui catchment. The total number of occurrences is 64, including all calibrations and validations for all the sub-periods.

Fig. 5 NSE values of the three models on the Lianshui catchment. Circles represent the NSE values of calibrations for every sub-period and the error bars are the NSE values of validations, with crosses indicating the mean. The performance scores of calibrations and validations exhibit a similar pattern.
occasionally cannot find the global optimal parameters. However, the differences are very small and therefore they are not considered to be significant.

4.1.2 The Wimmera catchment The results of RME are presented in Fig. 6 for the Wimmera catchment. The HBV model performs best in terms of RME, since the RME values concentrate around zero with small deviation. The XAJ model and the SimHYD model show biased RME values.

Figure 7 shows that the XAJ model gets the highest NSE values both in the calibrations and the validations, and the SimHYD model gets the lowest NSE values during the sub-periods P1 to P4. All models failed to give satisfactory simulation in the sub-period P5.

The ranges of NSE are the largest in the SimHYD model. The performance scores of the HBV model and the XAJ model are still very similar. The calibrated parameters on sub-period P5 also work on other sub-periods, whereas the calibrated parameters on other periods do not work on the sub-period P5. This can be seen from the NSE values of the HBV model and the XAJ model in Fig. 7. The sub-period P1 is also relatively dry, but model performances are different from the sub-period P5.

Compared with the Lianshui catchment, the performance scores of the three models are much lower and the ranges of NSE are larger on the Wimmera catchment. The reasons for model inefficiency are explained in Section 5. The larger ranges reveal larger parameter uncertainty than on the Lianshui catchment.

4.2 Stability of parameters in different calibration periods

The values of the optimized parameters in different sub-periods are an indicator of the stability of the parameters. The optimized parameter sets on the whole period are assumed to be most robust and most representative of catchment characteristics (Brigode et al. 2013), acting as a benchmark of parameters calibrated in the sub-periods. The changes of parameters are defined as the deviations in percent from their respective benchmark values, computed by:

\[ D_i = \left( \frac{P_i - P_w}{P_w} \right) \times 100, \quad P_w \neq 0 \]

where \( i \) is the index of the sub-period; \( P_i \) is the calibrated parameter value on the \( i \)th sub-period; \( P_w \) is the calibrated parameter value on the whole period.

The SUB parameter of the SimHYD model (Table 3)
calibrated on the whole period is zero, and hence this formula is not valid. However, the values of this parameter calibrated on all sub-periods on the two catchments are also zero. Therefore the deviations of SUB are zero. The cumulative frequencies of all the selected parameters (Tables 3, 4 and 5) within every 10% deviation intervals are shown in Fig. 8.

The comparisons between the two catchments indicate that the parameters of the three models are more stable on the Lianshui catchment than on the Wimmera catchment. On average, approximately 50% of the parameter values calibrated in the sub-periods are ±10% deviations from the benchmark values on the Lianshui catchment, whereas this number on the Wimmera catchment is only ~30%. Within ±50% deviation, cumulative frequency of parameters on Lianshui catchment can reach about 80–90%, whereas on the Wimmera catchment it is about 60–80%. Additionally, the ranks of the three models on the two catchments show a similar pattern. The differences between the models are smaller on the Lianshui catchment than on the Wimmera catchment.

For illustrative purposes, the parameter values are shown in Figs 9 and 10, for the Lianshui and

**Fig. 8** Cumulative frequency of deviations of parameters calibrated on all sub-periods on the Lianshui catchment (left) and the Wimmera catchment (right). The optimized parameters on the whole period are the benchmark values.

 calibrated on the whole period is zero, and hence this formula is not valid. However, the values of this parameter calibrated on all sub-periods on the two catchments are also zero. Therefore the deviations of SUB are zero. The cumulative frequencies of all the selected parameters (Tables 3, 4 and 5) within every 10% deviation intervals are shown in Fig. 8.

The comparisons between the two catchments indicate that the parameters of the three models are more stable on the Lianshui catchment than on the Wimmera catchment. On average, approximately 50% of the parameter values calibrated in the sub-periods are ±10% deviations from the benchmark values on the Lianshui catchment, whereas this number on the Wimmera catchment is only ~30%. Within ±50% deviation, cumulative frequency of parameters on Lianshui catchment can reach about 80–90%, whereas on the Wimmera catchment it is about 60–80%. Additionally, the ranks of the three models on the two catchments show a similar pattern. The differences between the models are smaller on the Lianshui catchment than on the Wimmera catchment.

For illustrative purposes, the parameter values are shown in Figs 9 and 10, for the Lianshui and
Wimmera catchments, respectively. They show that some parameters do not change at all and some are varying in different manners between the two catchments. For the SimHYD model, SUB keeps stable on both catchments during all sub-periods; CRAK, K and PC are within ±50% deviations on the Lianshui catchment during all sub-periods, whereas such parameter behaviour occurs for the parameters SMSC and K on the Wimmera catchment. For the HBV model, KUZ keeps stable on the Lianshui catchment during all sub-periods; PREC and EPO are within ±50% deviations on both catchments during all sub-periods. For the XAJ model, KE, SM, KG and NK are within ±50% deviations on the Lianshui catchment during all sub-periods, whereas such behaviour occurs for the parameters KE and KG on the Wimmera catchment. This means that the parameter values of the three models are more stable on the Lianshui catchment than on the Wimmera catchment.

5 DISCUSSION

The methods that split data series into various calibration and validation periods were put forward by Klemes (1986) for testing hydrological models under non-stationary climate conditions. Using a modified differential split-sample test, three widely used hydrological models, namely the SimHYD model, the HBV model and the XAJ model are evaluated on two catchments with changing climate conditions. On the Lianshui catchment, runoff and precipitation increased from 1989, whereas on the Wimmera catchment a severe drought started in 1997. All models performed better on the Lianshui catchment with a humid climate than on the Wimmera catchment with an arid climate. This confirmed that hydrological modelling on arid and semi-arid areas is much more difficult than on humid areas (Michaud and Sorooshian 1994, Zhang and Lindström 1996, Hughes 2008). In arid and semi-
arid areas, precipitation generally is characterized as short duration storms of high intensity. Associated with relatively thin vegetative cover and high evaporation rates, the dominant runoff generation mechanism is infiltration-excess overland flow. Rainfall and consequently runoff generation have large spatial variability, frequently occurring at local scales. This local scale character leads to some runoff, generated on some of the slopes, not always surviving to contribute to runoff at the outlet of catchments due to infiltration and evaporation.

However, it is also possible that water penetrates into deep groundwater aquifers, and finally contributes to runoff at the outlet of catchments (Ye et al. 1997). The complex spatial pattern and runoff generation mechanism would be a main reason for less model efficiency on the Wimmera catchment than on the Lianshui catchment.

If trading space for time, by these models, increasing precipitation would lead to higher model efficiency and vice versa. The two case studies clearly demonstrate this statement. On the Lianshui catchment, the NSE values of the three models increased during P6, P7 and P8 compared with P4 and P5 (Fig. 5). Conversely on the Wimmera catchment, the NSE values of three models decreased during the dry sub-period P5.

The stability either in model performance or parameter values obviously depends on how significant are the changes between the sub-periods. Less stable model performance and parameter values on the Wimmera catchment than on the Lianshui catchment are quite reasonable, since changes on the Wimmera catchment were more distinct than on the Lianshui catchment. Another possible reason is the uncertainty in the inputs and runoff data. As aforementioned, precipitation and runoff generation are highly localized and last for a short duration in arid and semi-arid areas. High variability both in space and time results in inadequate model presentation, either by lumping or from sparse resolution of in situ observations (Hughes 1995).

Data error or less efficiency is thought to be one of the possible reasons for parameter instability (Merz et al. 2011). Actually, at least one parameter is designed for data error in precipitation and potential evaporation in the three models. In the SimHYD model, PC and EC are multipliers respectively for precipitation and potential evaporation as tabulated in Table 3. On the Lianshui catchment, PC values were within ±50% deviations during eight sub-periods (in Fig. 9), whereas on the Wimmera catchment, calibrated PC values on four sub-periods were within ±50% deviations (in Fig. 10). Besides, on the Lianshui catchment, PC values were approximately one; whereas on the Wimmera catchment, PC values were negatively biased from one. Higher stable values of PREC in the HBV model and KE in the XAJ model on the Lianshui catchment than on the Wimmera catchment also substantiate this supposition.

Though all three models are conceptual hydrological models, some routines indeed provide insights for model development for arid and semi-arid areas. Both the HBV model and the XAJ model use exponential curves for the relationship between net rainfall and runoff (Bergström 1992, Zhao 1992, Zhang and Lindström 1996). This is very necessary and of high value to represent the localized runoff generation response to precipitation. However, the evaporation routine and separation of runoff components are more complex in the XAJ model than in the HBV model. Soil evaporation was calculated by dividing the soil into three layers in the XAJ model (Zhao 1992) whereas soil evaporation is only a nonlinear function of available water and soil capacity in the HBV model (Bergström 1992). The fine representation of soil evaporation is considered for improving evaporation modelling in a nonlinear soil system, especially on the Wimmera catchment. There are four runoff components in the XAJ model (Zhao 1992) and two in the HBV model (Lindström et al. 1997, Beldring et al. 2003).

6 CONCLUSIONS

Rainfall–runoff models are mathematical tools used to describe the relationship between precipitation and runoff. However, climate change significantly modifies this relationship. Hydrologists are required to evaluate if hydrological models can behave successfully on catchments with changing conditions. To contribute to such a study, we selected two catchments subject to changes in climate conditions, the Lianshui catchment in the humid zone in southern China and the Wimmera catchment in the arid zone in southeast Australia. We compared three conceptual models, the SimHYD model, the HBV model and the XAJ model in terms of stabilities of model performance and parameter values.

The study shows that the XAJ model gives the best model efficiency and the HBV model is very similar to the XAJ model in this respect. The SimHYD model has somewhat lower efficiency. All
models have a higher efficiency on the Lianshui catchment (wet) than on the Wimmera catchment (dry).

Performance scores as well as the stability of parameter values of all the models are more stable on the Lianshui catchment than on the Wimmera catchment. However, this conclusion is partly caused by the fact that runoff change on the Wimmera catchment is larger than on the Lianshui catchment.

Model structure determines model performance. Among the three models, the structures of the HBV model and the XAJ model are very similar and performance scores of the two models are also close on both catchments. The high performance scores of the models indicate that the structural assumptions are valid on the two catchments.

The stability of model performance and parameter values highly depends on the climate of catchments. The models with higher performance scores are more stable in changing conditions. This conclusion is drawn from the differences of individual model performance on two catchments. All the models are more stable on the Lianshui catchment than on the Wimmera catchment both in performance scores and parameter values.

For the parameter stability comparisons, the study shows that there is not much difference between the three models. There is not enough evidence to generalize which model parameter values are more stable although significant differences in performance scores exist among the models.

Acknowledgements The provision of data for the Lianshui catchment by Hua Chen in Wuhan University, China and for land use of the Wimmera catchment by Department of Geography, University of Maryland are gratefully acknowledged. The helpful comments by Guillaume Thirel in Irstea, France did motivate us to clarify our work and led to a substantial improvement of this paper. The authors are also grateful to the reviewers for their valuable comments and suggestions.

Disclosure statement No potential conflict of interest was reported by the author(s).

Funding This research contributes to the project-JOINTINDNOR (project number 203867) funded by the Research Council of Norway.

REFERENCES


Hofgaard, A., 1997. Inter-relationships between treeline position, species diversity, land use and climate change in the central


*Advances in Atmospheric Sciences*, 22 (6), 789–797. doi:10.1007/BF02918679.

Xu, H., et al., 2013. Assessing the influence of rain gauge density and distribution on hydrological model performance in a humid region of China. 


Zeng, S., Xia, J., and Du, H., 2014. Separating the effects of climate change and human activities on runoff over different time scales in the Zhang River basin. 


Zhao, R., 1992. The Xinanjiang model applied in China. 