Hydrol. Earth Syst. Sci. Discuss., 8, 8701–8736, 2011
www.hydrol-earth-syst-sci-discuss.net/8/8701/2011/
doi:10.5194/hessd-8-8701-2011

Hydrology and Earth System Sciences Discussions

The transferability of hydrological models under nonstationary climatic conditions

C. Z. Li[1,2], L. Zhang[2], H. Wang[1], Y. Q. Zhang[2], F. L. Yu[1], and D. H. Yan[1]

[1]State Key Laboratory of Simulation and Regulation of Water Cycle in River Basin, China Institute of Water Resources and Hydropower Research, Beijing, 100038, China
[2]CSIRO Land and Water, Canberra, ACT 2601, Australia

Received: 8 September 2011 – Accepted: 12 September 2011
– Published: 23 September 2011

Correspondence to: L. Zhang (lu.zhang@csiro.au)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

This paper investigates issues involved in calibrating hydrological models against observed data when the aim of the modelling is to predict future runoff under different climatic conditions. To achieve this objective, we tested two hydrological models, DWBM and SIMHYD, using data from 30 unimpaired catchments in Australia which had at least 60 years of daily precipitation, potential evapotranspiration (PET), and streamflow data. Nash-Sutcliffe efficiency (NSE) and absolute percentage water balance error (WBE) were used as performance criteria. We used a differential split-sample test to split up the data into 120 sub-periods and 4 different climatic sub-periods in order to assess how well the calibrated model could be transferred different periods. For each catchment, the models were calibrated for one sub-period and validated on the other three. Monte Carlo simulation was used to explore parameter stability compared to historic climatic variability. The chi-square test was used to measure the relationship between the distribution of the parameters and hydroclimatic variability. The results showed that the performance of the two hydrological models differed and depended on the model calibration. We found that if a hydrological model is set up to simulate runoff for a wet climate scenario then it should be calibrated on a wet segment of the historic record, and similarly a dry segment should be used for a dry climate scenario. The Monte Carlo simulation provides an effective and pragmatic approach to explore uncertainty and equifinality in hydrological model parameters. Some parameters of the hydrological models are shown to be significantly more sensitive to the choice of calibration periods. Our findings support the idea that when using conceptual hydrological models to assess future climate change impacts, a differential split-sample test and Monte Carlo simulation can reduce uncertainties due to parameter instability and non-uniqueness.

1 Introduction

Climate change caused by increasing atmospheric concentration of greenhouse gases may have significant effects on the hydrological cycle and water availability, affecting agriculture, forestry, and other industries and giving rise to significant economic consequences (Rind et al., 1992; IPCC, 2007). Changes in the hydrological cycle may indicate more floods and droughts, and increased pressure on water supply and irrigation systems. It is important to be able to estimate the potential impact of climate change on water resources and develop sustainable management strategies. One of the challenges is how to deal with hydrological nonstationarity (Milly et al., 2008). There are numerous factors that can affect hydrological stationarity and these include vegetation responses to elevated CO_2, changes in land use, and rainfall characteristics. It is crucial to improve our understanding of the effect of stationarity on hydrological assessments of climate change.

Hydrological models are important both for predicting the climate change scenarios and for assessing their hydrological and socioeconomic impacts. A variety of models have been used to evaluate hydrological effects (Rind et al., 1992). Predicting the hydrological impacts of climate change involves two key steps: downscaling the outputs from global climate models (GCMs) and then running hydrological models. At present, outputs from different GCMs have been used to drive hydrological models for predicting streamflow under a changed climate (Chiew et al., 2009). There are many factors that can affect the accuracy of a rainfall-runoff model in predicting the hydrological responses to climate change, and they include the particular hydrological model chosen, the GCM used, the optimisation technique employed, and the calibration period of the model. Most researchers usually use an ensemble of these techniques to minimise the uncertainty in predicting climate change impacts. For instance, Chiew et al. (1995) used results from 5 separate GCM experiments and reported that, in certain parts of Australia, the GCMs did not even agree on the direction of change in rainfall (i.e. increasing or decreasing rainfall). Boorman et al. (1997) used 2 conceptual

HESSD
8, 8701–8736, 2011

rainfall-runoff models in 3 catchments in UK and considered 2 climate scenarios and 8 climate sensitivity tests to quantify effects of climate change on 3 flow indices (mean runoff, flood magnitude, and low flow). Minville et al. (2008) produced an uncertainty envelope of future hydrological variables by considering 10 equally weighted climate projections from a combination of 5 GCMs and 2 greenhouse gas emission scenarios. Monomoy et al. (2007) used 6 automatic optimisation techniques for calibrating a conceptual rainfall-runoff model, and there have been a number of more recent studies estimating the impact of climate change on hydrological processes (Chiew et al., 2009; Vaze et al., 2010; Boyer et al., 2010). An implicit assumption in all these studies is that rainfall-runoff models calibrated over the historical period are valid for predicting the future hydrological regime under a changed climate and this relates directly to the assumption of hydrological stationarity. However, little has been carried out to test the validity of this assumption.

Calibration of hydrological models generally involves optimizing model parameters to match measured streamflow using observed rainfall as input. For predicting the impact of climate change, the input rainfall series are varied according to an assumed future climate scenario. But is it appropriate to use these models under future climatic conditions when rainfall-runoff relations could be very different to those experienced historically?

This paper investigates the transferability of hydrological models under nonstationary climatic conditions. We compare results obtained with different hydrological models calibrated under different climatic conditions. The paper first presents two hydrological models chosen for this study – the Dynamic Water Balance Model (DWBM) and the SIMHYD model – and then describes the data used to calibrate them. We describe different methods of applying the data, including a split-sample test, a Monte Carlo simulation, and a performance criterion. Finally, we analyse the performance of the models under different calibration conditions and discuss the optimal parameters for each.

HESSD
8, 8701–8736, 2011

2 Description of hydrological models and data

Two lumped hydrological models with daily inputs were chosen for this study: the Dynamic Water Balance Model (DWBM) (Zhang et al., 2008) and the SIMHYD model (Chiew et al., 2002), and detailed description of the two models is presented below.

2.1 The Dynamic Water Balance Model (DWBM)

The DWBM model used in this study was developed by Zhang et al. (2008). It is a lumped conceptual water balance model with two stores: a near surface root-zone store and a deeper zone store (Fig. 1). The model is based on Budyko's concept of water availability and atmospheric demand (Budyko, 1958) or the concept of "*limits and controls*" (Calder, 1998). Fundamental to this model is a functional form that represents a smooth transition between supply and demand limits (Fu, 1981):

$$\frac{E}{P} = 1 + \frac{E_0}{P} - \left[1 + \left(\frac{E_0}{P}\right)^w\right]^{1/w} \tag{1}$$

where w is a model parameter ranging between 1 and ∞. For the purpose of model calibration, we define $\alpha = 1 - 1/w$ so that α varies between 0 and 1. This definition also conveniently associates an increase in α with an increase in evapotranspiration efficiency. P is rainfall and E_0 is potential evapotranspiration at mean annual timescale. More details of this mean annual water balance model are given in Zhang et al. (2004, 2008).

It is assumed that rainfall $P(t)$ in time step t will be partitioned into direct runoff $Q_d(t)$ and catchment rainfall retention:

$$P(t) = Q_d(t) + X(t) \tag{2}$$

where $X(t)$is called catchment rainfall retention and is the amount of rainfall retained by the catchment for evapotranspiration ET(t), change in soil moisture storage $S(t) - S(t-1)$ and recharge $R(t)$.

HESSD
8, 8701–8736, 2011

The demand limit for $X(t)$ is the sum of available storage capacity ($S_{max} - S(t-1)$) and potential evapotranspiration ($E_0(t)$) and is denoted as $X_0(t)$, while the supply limit can be considered as rainfall $P(t)$. Following a similar argument to Budyko (1958), we can postulate that:

$$X(t)/P(t) \rightarrow 1 \quad \text{as} \quad X_0(t)/P(t) \rightarrow \infty \quad \text{(very dry conditions)} \tag{3}$$

$$X(t) \rightarrow X_0(t) \quad \text{as} \quad X_0(t)/P(t) \rightarrow 0 \quad \text{(very wet conditions)}. \tag{4}$$

The catchment rainfall retention $X(t)$ can be calculated as:

$$X(t) = P(t)\,F\left(\frac{X_0(t)}{P(t)}, \alpha_1\right) \tag{5}$$

where $F()$ is Fu's curve – Eq. (1), α_1 is rainfall retention efficiency, i.e. a larger α_1 value will result in more rainfall retention and less direct runoff.

From Eqs. (1) and (5), direct runoff is calculated as:

$$Q_d(t) = P(t) - X(t). \tag{6}$$

At sub-annual time scales, water availability $W(t)$ can be defined as:

$$W(t) = X(t) + S(t-1). \tag{7}$$

Combining the definition of $X(t)$ with Eq. (7), one obtains:

$$W(t) = \mathrm{ET}(t) + S(t) + R(t). \tag{8}$$

While Eq. (7) defines the source of the water availability, Eq. (8) determines the partitioning. Next define evapotranspiration opportunity (Sankarasubramanian and Vogel, 2002) as $Y(t) = \mathrm{ET}(t) + S(t)$, we obtain:

$$W(t) = Y(t) + R(t). \tag{9}$$

HESSD
8, 8701–8736, 2011

The demand limit for $Y(t)$ can be considered as the sum of potential evapotranspiration ($E_0(t)$) and soil water storage capacity (S_{max}) and is denoted as $Y_0(t)$, while the supply limit is the available water $W(t)$. Similar to Budyko (1958), we can postulate that:

$$Y(t)/W(t) \rightarrow 1 \quad \text{as } Y_0(t)/W(t) \rightarrow \infty \quad \text{(very dry conditions)} \tag{10}$$

$$Y(t) \rightarrow Y_0(t) \quad \text{as} \quad Y_0(t)/W(t) \rightarrow 0 \quad \text{(very wet conditions)}. \tag{11}$$

The evapotranspiration opportunity $Y(t)$ can be estimated from the following relationship:

$$Y(t) = W(t)\, F\left(\frac{E_0(t) + S_{max}}{W(t)}, \alpha_2\right). \tag{12}$$

Thus groundwater recharge $R(t)$ can be calculated from Eq. (9). The next step is to calculate evapotranspiration $\mathrm{ET}(t)$. The demand limit for $\mathrm{ET}(t)$ can be considered as potential evapotranspiration $E_0(t)$ and the supply limit is the available water $W(t)$. Similar to Budyko (1958), evapotranspiration $\mathrm{ET}(t)$ can be calculated as:

$$\mathrm{ET}(t) = W(t)\, F\left(\frac{E_0(t)}{W(t)}, \alpha_2\right) \tag{13}$$

where α_2 is a model parameter, representing evapotranspiration efficiency.

Soil water storage can now be calculated as:

$$S(t) = Y(t) - \mathrm{ET}(t). \tag{14}$$

Finally, groundwater storage is treated as linear reservoir, so that baseflow and groundwater balance can be modelled as:

$$Q_b(t) = dG(t-1) \tag{15}$$

$$G(t) = (1-d)\, G(t-1) + R(t) \tag{16}$$

where Q_b is baseflow, G is groundwater storage, and d is a recession constant.

HESSD
8, 8701–8736, 2011

The DWBM model has been applied to 265 catchments in Australia and showed encouraging results (Zhang et al., 2008). The model has four parameters: retention efficiency (α_1); evapotranspiration efficiency (α_2); soil water storage capacity (S_{max}), and baseflow linear recession constant (d). The range of the parameter values is shown in Table 1.

2.2 The SIMHYD model

The SIMHYD model is a lumped conceptual daily rainfall-runoff model (Chiew et al., 2002), driven by daily rainfall and PET, which simulates daily streamflow. It has been tested and used extensively across Australia (Chiew et al., 2002; Siriwardena et al., 2006; Viney et al., 2008; Zhang et al., 2008, 2009). Figure 2 shows the structure of the SIMHYD model and the algorithms controlling how water enters the system from precipitation, flows into several stores, and then flows out through evapotranspiration and runoff. The SIMHYD model has 7 parameters, and the useful ranges of them are shown in Table 2.

In the SIMHYD model, daily rainfall is first intercepted by an interception store, which is emptied each day by evaporation. Incident rainfall, which occurs if rainfall exceeds the maximum daily interception, is then subjected to an infiltration function. The incident rainfall that exceeds the infiltration capacity becomes infiltration excess runoff. A soil moisture function diverts the infiltrated water to the river (as saturation excess runoff/interflow), groundwater store (as recharge) and soil moisture store. The saturation excess runoff/interflow is first estimated as a linear function of the soil wetness (soil moisture level divided by soil moisture capacity). The equation used to simulate interflow therefore attempts to mimic both the interflow and saturation excess runoff processes (with soil wetness used to reflect those parts of the catchment that are saturated and from which saturation excess runoff can occur). Groundwater recharge is then estimated, also as a linear function of the soil wetness. The remaining moisture flows into the soil moisture store. Evapotranspiration from the soil moisture store is estimated as a linear function of the soil wetness, but cannot exceed the potential

HESSD
8, 8701–8736, 2011

rate (PET minus intercepted water). The soil moisture store has a finite capacity and overflows into the groundwater store, baseflow from which is simulated as a linear recession from the groundwater store. The model has therefore three runoff components: infiltration excess runoff, saturation excess runoff/interflow, and baseflow.

2.3 Study catchments and data

In this study 30 catchments from Australia were selected with at least 60 years of unimpaired daily streamflow data (Fig. 3). Unimpaired streamflow is defined as streamflow that is not subject to regulation or diversion. The catchment area ranges from 82 to 1891 km^2 with mean annual streamflow varied between 53 to 1363 mm. The mean annual precipitation (P) ranges from 628 to 2095 mm and annual potential evapotranspiration (PET) ranges from 817 to 2098 mm, representing diverse hydrological and climatic conditions. The runoff coefficient varies from 0.08 to 0.65.

Catchment averaged annual rainfall was estimated from gridded SILO daily rainfall (http://www.longpaddock.qld.gov.au/silo, Jeffrey et al., 2001). The SILO Data Drill provides surfaces of daily rainfall and other climate data interpolated from point measurements made by the Australian Bureau of Meteorology. The spatial resolution of the gridded daily rainfall data is 0.05° based on interpolation of over 6000 rainfall stations across Australia. The interpolation uses monthly rainfall data, ordinary kriging with zero nugget, and a variable range. Monthly rainfall for each 5×5 km grid cell was converted to daily rainfall using daily rainfall distribution from the station closest to the grid cell (Jeffrey et al., 2001). The daily time series of maximum and minimum temperatures, incoming solar radiation, actual vapour pressure, and precipitation at 0.05×0.05 (~5 km × 5 km) grid cells from the SILO Data Drill (http://www.longpaddock.qld.gov.au/silo) were used.

Potential evaporation was alculated using the Priestley-Taylor equation (Priestley and Taylor, 1972) for each catchment with the Priestley-Taylor coefficient set to 1.26 following Raupach (2000). In the calculation, the available energy was taken as equal to the net radiation by neglecting ground heat flux. The net radiation was calculated from the

HESSD
8, 8701–8736, 2011

incoming global shortwave and longwave radiation, surface albedo, surface emissivity, and surface temperature as described by Raupach et al. (2001).

Daily streamflow data were obtained from the Australian Land and Water Resources Audit project (Peel et al., 2000) and have been quality checked. Firstly, data quality codes were checked for any missing and poor-quality data as most gauging stations provide numerical codes indicating quality of streamflow data. Missing streamflow data were infilled by interpolating streamflow values at previous and following days. Secondly, time series of daily rainfall and streamflow were plotted to identify any inconsistency and recording errors in the data (e.g. spikes, same streamflow value for a long period of time). The quality checks are to ensure good quality streamflow data are used in the study.

3 Methods

3.1 Differential split-sample test

In general, hydrological models rely on stationary conditions (Xu, 1999). Usually, model calibration requires a split-sample test, where the model is calibrated during one climatic period and validated on another independent period. The split-sample test is the classical test, being applicable to cases where there is sufficiently long time series of the climatic data for both calibration and validation and where the catchment conditions remain unchanged, i.e. stationary (Refsgaard and Storm, 1996). This test gives an indication how the model might perform for an independent period having similar conditions. Unfortunately, this test is unable to guarantee the applicability of hydrological models under nonstationary conditions (Xu, 1999; Henriksen et al., 2003).

In order to try to answer the question of whether the transfer of parameter values from the present-day climate to a future climate is justified, the “differential split-sample test” proposed by Klemes (1986) was considered, in which the hydrological model is

HESSD
8, 8701–8736, 2011

tested on calibration and validation periods under contrasting climatic conditions. In this case, different sub-periods are chosen with different historical rainfall conditions.

In this study, different periods with various climatic conditions were identified. First of all, we calculated annual and mean annual precipitation over the whole period of record for each catchment. Then sub-periods with consecutive annual precipitation greater than the mean were selected as the "wet" periods and sub-periods with consecutive annual precipitation less than the mean were selected as the "dry" periods. The average annual precipitation for the "wet" and "dry" periods ranges from 10.2 % to 47.1 % and −10.4 % to −28.3 % of the long-term average annual precipitation, respectively. In the selection, the minimum length of the sub-period was set to 5 years to ensure stable model calibration. If this process results in more than two "wet" or "dry" periods, then the two wettest periods or two driest periods were selected for model calibration and validation (Fig. 4). The hydrological model was calibrated for each of the 4 sub-periods and validated on each of the remaining 3 sub-periods in turn, resulting in a total of 12 calibration and validation tests.

To examine model performance under different calibration and validation conditions, results from the above tests are grouped as "*dry*/dry", "*dry*/wet", "*wet*/wet", and "*wet*/dry" to represent climatic conditions in the calibration and validation periods respectively.

3.2 Monte Carlo simulation

The concept of equifinality has become widely recognised during recent years (Beven, 1993; Boorman et al., 1997; Niel et al., 2003; Wilby et al., 2005; Minville et al., 2008). Often parameter sets, which perform equally well for a calibration period, can be found at very different locations in the parameter space. It may be argued that the problem of identifying a unique parameter set is not an issue for practical model applications, i.e. if different parameter sets were equally suitable to simulate runoff during a calibration period, any one of these parameter sets may be applied. However, these equally good parameter sets may give different predictions when the model is used to estimate the effects of land use and climate change on streamflow (Uhlenbrook et al., 1999). The need for improved model calibration and testing has been emphasized in recent years. Monte Carlo simulation is an effective way of calculating confidence limits of predicted time series and exploring parameter stability and identifiability in the context of historic climate variability (Uhlenbrook et al., 1999; Wilby, 2005; Widen-Nilsson et al., 2009).

In this study, Monte Carlo simulation was undertaken with 1 000 000 runs, each with a randomly generated parameter set for the calibration period and the lower and upper limits of the parameters are listed in Tables 1 and 2. We then selected assemblies of the 100 best parameter sets according to a goodness-of-fit measure which is defined in Sect. 3.3. Finally, the models were run during the validation periods with all the best parameter sets. Calibration with the 100 best parameter sets gave very similar results, and to minimise fluctuations, the average results were used for subsequent analysis.

3.3 Assessment model performance criteria

The Nash-Sutcliffe efficiency (NSE) (Nash and Sutcliffe, 1970) was used as the statistic criterion of the model performance. The models were calibrated to maximize the Nash and Sutcliffe efficiency of daily runoff, which is defined as:

$$\mathrm{NSE} = 1 - \frac{\sum_{i=1}^{N} \left(Q_{\mathrm{obs},i} - Q_{\mathrm{sim},i}\right)^2}{\sum_{i=1}^{N} \left(Q_{\mathrm{obs},i} - \overline{Q}_{\mathrm{obs},i}\right)^2} \tag{17}$$

where $Q_{\mathrm{sim},i}$ and $Q_{\mathrm{obs},i}$ are the simulated and observed daily runoff, respectively, $\overline{Q}_{\mathrm{obs},i}$ is the mean observed runoff, i is the i-th day, and N is the number of days sampled and it varies with individual catchment.

As well as NSE of daily runoff, we used another criterion – absolute percentage water balance error (WBE) – to measure the model performance during the calibration

HESSD
8, 8701–8736, 2011

and validation periods. WBE is a measure of the bias in the validation results from the observed flow (Hogue et al., 2006) and absolute WBE is defined as:

$$\text{absolute WBE} = \frac{\sum_{i=1}^{N} \left| Q_{\text{sim},i} - Q_{\text{obs},i} \right|}{\sum_{i=1}^{N} Q_{\text{obs},i}} \times 100\,\% \tag{18}$$

with the symbols defined above.

3.4 Analysis of parameter probability distributions under different calibration periods

For each of the models, we ended up with 100 best parameter sets for each catchment and for each calibration period. From these parameters sets we could then calculate a probability distribution of each parameter. For a given significance level α, the chi-square test (χ^2 test) was used to test the null hypothesis that the parameter distributions obtained for a dry period and a wet period were significantly different. A p value greater than 0.01 indicates we can reject the null hypothesis, which also means that the parameter probability distributions for the two different calibration periods are similar.

4 Results and discussion

4.1 Comparisons of model calibration under different climatic conditions

Results of model calibration under different climatic conditions are shown in Figs. 5 and 6 and Table 3. Figure 5a shows the percentage of model calibration tests that have a NSE value exceeding a given value, and Fig. 5b is a corresponding plot for absolute WBE values. It can be seen that the SIMHYD model was well calibrated under both dry and wet conditions with average NSE value greater than 0.7. The average water

HESSD
8, 8701–8736, 2011

balance error is 14 % and 11 % for the dry and wet calibration periods. Compared with the SIMHYD model, the DWBM model showed slightly poor results in terms of NSE and absolute WBE. The plots show that both models are better calibrated under wet periods than under dry ones, with higher NSE values and lower absolute WBE values. For example, under the dry conditions, average NSE was 0.7 and 0.57 for the SIMHYD and the DWBM model. Under the wet conditions, average NSE was 0.76 and 0.65 respectively for the two models. In Fig. 5a, better performance means a larger area under the curve, whereas in Fig. 5b, a smaller area is better. All the results became worse when the calibration periods became drier, indicating a higher sensitivity of the models to dry climatic conditions. The results also indicated that the errors in the simulated runoff were increased under drier climatic conditions.

Figure 6 presents the same results as box and whisker plots of the distributions of the NSE and absolute WBE values. Under dry and wet calibration periods, the median percentile of the NSE values from the 60 tests are, for the SIMHYD model, 0.70 and 0.77, respectively, and for the DWBM model, 0.58 and 0.66. The median percentile of the absolute WBE values from the 60 tests are 13 % and 8 % for the SIMHYD model under dry and wet calibration periods respectively, and 15 % and 12 % for the DWBM model. These results indicate that the two models can be calibrated satisfactorily for most of the tests, although the calibration results of the DWBM model are slightly poorer compared with those of the SIMHYD model. The average NSE values calibrated under the wet periods are higher – i.e. better – by 0.06 (SIMHYD model) and 0.08 (DWBM model) than those calibrated under dry periods. The average absolute WBE values calibrated under wet periods are lower – again better – by 3 % (SIMHYD model) and 5 % (DWBM model) than those calibrated under the dry period.

Streamflow of a catchment is influenced by a number of factors, most noticeably rainfall and antecedent soil moisture. During dry periods, catchments are generally characterized by small runoff events and lower runoff to rainfall ratios with higher percentage error in both rainfall and runoff. In this case, rainfall-runoff models become very sensitive to both rainfall and parameter optimization. Also, dry periods may not

HESSD
8, 8701–8736, 2011

contain enough high flows to adequately calibrate model parameters responsible for simulating high flows (Gan et al., 1997). Apart from rainfall amount, spatial variability of rainfall can also affect runoff. Smith et al. (2004) showed that improved runoff simulations can be obtained from distributed versus lumped rainfall-runoff models in catchments with considerable rainfall variability. Spatial variability of rainfall was also found to be the dominant control on runoff production (Segond et al., 2007). In this study, spatially averaged rainfall was used in both model calibration and validation. This is likely to affect the model results and it is expected that the rainfall variability effect will be greater in dry periods than in wet periods. It has been widely acknowledged that spatial variability of antecedent soil moisture conditions plays an important role in runoff generation (Grayson and Blöschl, 2000). Minet et al. (2011) investigated the effect of spatial soil moisture variability on runoff simulations using a distributed hydrologic model and showed that model results are sensitive to soil moisture spatial variability, especially in dry conditions. At catchment scales, soil moisture exhibit larger heterogeneity under dry conditions than wet conditions and this means errors associated with dry period runoff simulations are likely to be greater as runoff generation exhibits non-linear threshold behaviour.

4.2 Comparisons of model validation using different calibration periods

Validation runs were conducted for 60, 120, 60, and 120 tests for the *dry*/dry, *dry*/wet, *wet*/dry, and *wet*/wet groups respectively. The model validation results are summarized in Table 4 and Figs. 7 and 8. As expected, the validation results are slightly poorer than the calibration results, with the averaged NSE values in the model validation generally being 0.1 to 0.2 lower than those in the model calibration. Table 4 summarizes the 25th percentile, median, 75th percentile, and average of NSE values and absolute WBE values in the validation periods. As Table 4 indicates, average values differed little between calibration periods, although Figs. 7 and 8 show some interesting features.

Comparing the validation results of the *dry*/dry, *dry*/wet, *wet*/dry, and *wet*/wet groups in Fig. 7a–d, it can be noted both the SIMHYD and DWBM models gave similar

HESSD
8, 8701–8736, 2011

The transferability of hydrological models under nonstationary climatic conditions

C. Z. Li et al.

patterns. The results for the *wet*/wet are better than those of the *dry*/wet – this means that the models performed better during a wet period when they are calibrated against a wet period, compared to when they are calibrated against a dry period. These results suggest, not unexpectedly, that if a hydrological model is intended to simulate streamflow for a wet climate period then it should be calibrated on a wet segment of the historic record. They also show that hydrological models will, in general, perform better when calibrated in a wet period than when calibrated in the dry period.

As is shown in Table 4 and Fig. 8, the results from the *dry*/dry test are slightly better than the results from the *wet*/dry test. The results indicate, again reasonably, that the hydrological models perform better in a dry period when calibrated in a dry period rather than in a wet period. Similar results have been reported by Vaze et al. (2010). A closer examination of model errors reveals that when the model parameters, calibrated on a dry period, were used to simulate runoff during a wet period, the mean of the simulated runoff was usually underestimated; conversely, when model parameters, calibrated on a wet period, were used to simulate dry period runoff, the mean simulated runoff was overestimated, consistent with the findings of Gan et al. (1997). The differences in average annual rainfall between the wet and dry periods ranged from 10 to 47 % and are comparable with percentage change in man annual rainfall for 2030 relative to 1990 from 15 GCMs for the Murray Darling Basin in Australia (Chiew et al., 2008).

4.3 Parameter uncertainty analysis under climatic nonstationarity

As described in Sect. 3.2, assemblies of the 100 best parameter sets were selected from Monte Carlo simulation under different calibration conditions. Table 5 shows the percent of the catchments in which the model parameter distributions for a dry and wet period were significantly different ($p < 0.01$) using Monte Carlo simulation. For each model, the model parameters are ranked from the most sensitive to calibration conditions to least sensitive. For the SIMHYD model, the most sensitive parameters were SUB, SMSC, SQ, and CRAK, each of which significantly affected 50 % or more of the catchments. The other three parameters, K, COEFF, and INSC had smaller effects,

HESSD
8, 8701–8736, 2011

with INSC (having an effect in only 10 % of catchments) being the most insensitive to choice of dry and wet calibration periods.

In order to examine the effects of climatic conditions on the results, we grouped the 30 study catchments into two climatic types: 16 water-limited catchments with an index of dryness (E_p/P) greater than 1, and 14 energy-limited catchments with an index of dryness less than 1. It can be noted that all parameters performed differently in water-limited and energy-limited catchments, in particular SUB, SMSC, and CRAK. As described in Sect. 2.1, SUB (used in the estimation of interflow), SMSC (soil moisture store capacity), and CRAK (used in the estimation of groundwater recharge) are all soil related parameters. These results suggest that the soil parameters are very sensitive to climatic conditions.

For the DWBM model, the parameters α_1 and $S_{\max}$ exhibited different effects on runoff under the dry and wet calibration periods as 67 % and 63 % of the catchments showed statistically different results at the 0.01 level. At the other extreme, the parameter α_2 displayed an apparent insensitivity to the calibration periods (just 23% of catchments were affected). The parameter α_2 represents evapotranspiration efficiency and it behaves similarly to the parameter w of Zhang et al. (2001) and (2004), which was shown to be mostly correlated with vegetation cover. The parameter d was more sensitive to the choice of the calibration period for the water-limited catchments than for the energy-limited catchments. It is interesting to note that all the parameters behaved differently under the water-limited and energy-limited conditions, except perhaps for parameter α_2.

The results indicate that calibration periods can cause significant shifts in model parameter distributions. Some model parameters are relatively sensitive to the choice of calibration periods, while other parameters are fairly insensitive. As well as the impact of calibration periods on parameter distributions, whether catchments are water-limited or energy-limited also needs to be taken into consideration. These findings have major implications for studies of climate change impact on streamflow. When a hydrological model calibrated for a given climatic condition (e.g. wet periods) is used to simulate runoff of different climatic conditions (e.g. dry periods), transfer of some model parameters (i.e. sensitive parameters) may result in large errors in simulated runoff. Monte Carlo simulation provided an effective and pragmatic approach to exploring uncertainty and equifinality in hydrological model parameters. Rainfall-runoff models simulate hydrological processes and their performance is related to catchment characteristics such as climate, topography, soil, vegetation, catchment shape, geology, drainage network. In such a complex situation, it is hard to pinpoint the source of parameter uncertainty, but our results show that calibration periods and catchment climatic conditions are both important factors that can produce uncertainty in model performance.

5 Conclusions

Potentially large uncertainties arise when predicting hydrological responses to future climate change – due to factors such as the choice of emission scenario, GCM, downscaling technique, hydrological model, optimization technique, and the way the model is calibrated. It is therefore important to develop reliable ways to calibrate hydrological models under present-day conditions. This study compared hydrological model performances under nonstationarity by using the differential split-sample test and two conceptual rainfall-runoff models, DWBM and SIMHYD, applied to 30 catchments in Australia. Monte Carlo simulation was used to explore parameter stability and transferability in the context of historic climate variability.

Hydrological models differ in performance depending on how they are calibrated. If a hydrological model is intended to simulate runoff for a wet climate scenario then it should be calibrated on a wet segment of the historic record. Conversely, if it is intended to simulate runoff for a dry climate scenario then it should be calibrated on a dry segment of the historic record. Therefore, careful selection of the calibration period can reduce the modelling uncertainty when exploring future climate scenarios.

For both our models we found that the "*dry*/wet" tests performed better – had higher NSE values and lower absolute WBE values – than the "*wet*/dry" tests. In other words,

HESSD
8, 8701–8736, 2011

transferability of model parameter values from dry periods to wet periods is greater than vice versa, perhaps because of the more uniform rainfall and soil moisture conditions in the wet periods (Gan et al., 1997).

The choice of calibration period is a key step in predicting the impact of climate change on runoff. Our research has implications for hydrological modellers looking to estimate future runoff and we hope this study will stimulate further research into the selection of calibration data.

Acknowledgements. This study was supported by the National Basic Research Program of China (2010CB951102), the Foundation for Innovative Research Groups of the National Natural Science Foundation of China (51027006) and the Regional Water Theme in the Water for a Healthy Country Flagship. Lu Zhang acknowledges the Chinese Academy of Sciences for support through their Outstanding Overseas Chinese Scholars Fund. We thank Andrew Bell and Enli Wang for their helpful comments on a draft of the paper.

References

Beven, K. J.: Prophecy, reality and uncertainty in distributed hydrological modelling, Adv. Water Resour., 16, 41–51, 1993.

Boorman, D. B. and Sefton, C. E. M.: Recognising the uncertainty in the quantification of the effects of climate change on hydrological response, Climatic Change, 35, 415–434, 1997.

Boyer, C., Chaumont, D., Chartier, I., and Roy, A. G.: Impact of climate change on the hydrology of St. Lawrence tributaries, J. Hydrol., 384, 65–83, 2010.

Budyko, M. I.: The Heat Balance of the Earth's Surface, US Department of Commerce, Washington, DC, 1958.

Calder, I. R.: Water use by forests, limits and controls, Tree Physiol., 18, 625–631, 1998.

Chiew, F. H. S. and McMahon, T. A.: Global ENSO-streamflow teleconnection, streamflow forecasting and interannual variability, Hydrolog. Sci. J. – Journal Des Sciences Hydrologiques, 47, 505–522, 2002.

Chiew, F. H. S., Whetton, P. H., McMahon, T. A., and Pittock, A. B.: Simulation of the impacts of climate change on runoff and soil moisture in Australian catchments, J. Hydrol., 167, 121–147, 1995.

HESSD

8, 8701–8736, 2011

The transferability of hydrological models under nonstationary climatic conditions

C. Z. Li et al.

Chiew, F. H. S., Peel, M. C., and Western, A. W.: Application and testing of the simple rainfall-runoff model SIMHYD, in: Mathematical Models of Small Watershed Hydrology and Applications, edited by: Singh, V. P. and Frevert, D. K., Water Resources Publication, Littleton, Colorado, USA, 335–367, 2002.
Chiew, F. H. S., Teng, J., Kirono, D., Frost, A. J., Bathols, J. M., Vaze, J., Viney, N. R., Young, W. J., Hennessy, K. J., and Cai, W. J.: Climate data for hydrologic scenario modelling across the Murray-Darling Basin, A report to the Australian Government from the CSIRO Murray-Darling Basin Sustainable Yields Project, Water for a Healthy Country Flagship, CSIRO, 42 pp., 2008.
Chiew, F. H. S., Teng, J., Vaze, J., Post, D. A., Perraud, J. M., Kirono, D. G. C., and Viney, N. R.: Estimating climate change impact on runoff across southeast Australia: method, results, and implications of the modeling method, Water Resour. Res., 45, W10414, doi:10.1029/2008WR007338, 2009.
Fu, B. P.: On the calculation of the evaporation from land surface, Sci. Atmos. Sin., 23–31, 1981.
Gan, T. Y., Dlamini, E. M., and Biftu, G. F.: Effects of model complexity and structure, data quality, and objective functions on hydrologic modelling, J. Hydrol., 192, 81–103, 1997.
Grayson, R. B. and Blöschl, G.: Spatial Patterns in Catchment Hydrology: Observations and Modelling, Cambridge University Press, 404 pp., 2000.
Henriksen, H. J., Troldborg, L., Nyegaard, P., Sonnenborg, T. O., Refsgaard, J. C., and Madsen, B.: Methodology for construction, calibration and validation of a national hydrological model for Denmark, J. Hydrol., 280, 52–71, 2003.
Hogue, S. T., Gupta, H., and Sorooshian, S.: A "user-friendly" approach to parameter estimation in hydrologic models, J. Hydrol., 320, 202–217, 2006.
IPCC: Climate Change 2007: The Physical Basis, Contributions of Working Group 1 to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M. and Miller, H. L., Cambridge University Press, Cambridge, UK and New York, USA, 996 pp., 2007.
Jeffrey, S. J., Carter, J. O., Moodie, K. B., and Beswick, A. R.: Using spatial interpolation to construct a comprehensive archive of Australian climate data, Environ. Modell. Softw., 16, 309–330, 2001.
Klemes, V.: Operational testing of hydrological simulation models, Hydrolog. Sci. J., 31, 13–24, 1986.

HESSD
8, 8701–8736, 2011

The transferability of hydrological models under nonstationary climatic conditions

C. Z. Li et al.

Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P., and Stouffer, R. J.: Stationarity is dead: whither water management?, Science, 319, 573–574, 2008.
Minet, J., Laloy, E., Lambot, S., and Vanclooster, M.: Effect of high-resolution spatial soil moisture variability on simulated runoff response using a distributed hydrologic model, Hydrol. Earth Syst. Sci., 15, 1323–1338, doi:10.5194/hess-15-1323-2011, 2011.
Minville, M., Brissette, F., and Leconte, R.: Uncertainty of the impact of climate change on the hydrology of a nordic watershed, J. Hydrol., 358, 70–83, 2008.
Monomoy, Y. G. and Kieran, M. C.: Comparative assessment of six automatic optimization techniques for calibration of a conceptual rainfall-runoff model, Hydrolog. Sci. J. – Journal Des Sciences Hydrologiques, 52, 432–449, 2007.
Nash, J. E. and Sutcliffe, J. V.: River forcasting using conceptual models, 1. A discussion of principles, J. Hydrol., 10, 280–290, 1970.
Niel, H., Paturel, J. E., and Servat, E.: Study of parameter stability of a lumped hydrologic model in a context of climatic variability, J. Hydrol., 278, 213–230, 2003.
Peel, M. C., Chiew, F. H. S., Western, A. W., and McMahon, T. A.: Extension of unimpaired monthly stream flow data and regionalization of parameter values to estimate stream flow in ungauged catchments, Report to National Land and Water Resources Audit, Cent. For Environ. Appl. Hydrol., Univ. of Melbourne, Parkville, Vic., Australia, 2000.
Priestley, C. H. B. and Taylor, R. J.: On the assessment of the surface heat flux and evaporation using large-scale parameters, Mon. Weather Rev., 100, 81–92, 1972.
Raupach, M. R.: Equilibrium evaporation and the convective boundary layer, Bound.-Lay. Meteorol., 96, 107–141, 2000.
Raupach, M. R., Kirby, J. M., Barrett, D. J., Briggs, P. R., Lu, H., and Zhang, L.: Balances of water, carbon, nitrogen and phosphorus in Australian landscapes: 2. Model formulation and testing, Tech. Rep. 41/01, CSRIO Land and Water, Canberra, ACT, Australia, 2001.
Refsgaard, J. C. and Storm, B.: Construction, calibration and validation of hydrological models, in: Distributed Hydrological Modelling, edited by: Abbott, M. B. and Refsgaard, J. C., Kluwer Academic Publishers, The Netherlands, 50 pp., 1996.
Rind, D., Rosenzweig, C., and Goldberg, R.: Modelling the hydrological cycle in assessments of climate change, Nature, 358, 119–120, 1992.
Sankarasubramanian, A. and Vogel, R. V.: Hydroclimatology of the continental United tates, Geophys. Res. Lett., 30, 1363, doi:10.1029/2002GL015937, 2003.

HESSD
8, 8701–8736, 2011

The transferability of hydrological models under nonstationary climatic conditions

C. Z. Li et al.

Segond, M. L., Wheater, H. S., and Onof, C.: The significance of spatial rainfall representation for flood runoff estimation: A numerical evaluation based on the Lee catchment, UK, J. Hydrol., 347, 116–131, 2007.
Siriwardena, L., Finlayson, B. L., and McMahon, T. A.: The impact of land use change on catchment hydrology in large catchments: The Comet River, Central Queensland, Australia, J. Hydrol., 326, 199–214, 2006.
Smith, M. B., Koren, V. I., Zhang, Z. Y., Reed, S. M., Pan, J. J., and Moreda, F.: Runoff response to spatial variability in precipitation: an analysis of observed data, J. Hydrol., 298, 267–286, 2004.
Uhlenbrook, S., Seibert, J., Leibundgut, C., and Rodhe, A.: Prediction uncertainty of conceptual rainfall-runoff models caused by problems in identifying model parameters and structure, Hydrolog. Sci. J. – Journal des Sciences Hydrologiques, 44, 779–797, 1999.
Vaze, J., Post, D. A., Chiew, F. H. S., Perraud, J. M., Viney, N. R., and Teng, J.: Climate non-stationarity – Validity of calibrated rainfall-runoff models for use in climate change studies, J. Hydrol., 394, 447–457, 2010.
Viney, N., Vaze, J., Chiew, F., and Perraud, J.: Regionalisation of runoff generation across the Murray-Darling Basin using an ensemble of two rainfall-runoff models, Paper presented at Water Down Under 2008, April 2008, Adelaide, Engineers Australia, 2008.
Widen-Nilsson, E., Gong, L., Halldin, S., and Xu, C. Y.: Model performance and parameter behavior for varying time aggregations and evaluation criteria in the WASMOD-M global water balance model, Water Resour. Res., 45, W05418, doi:10.1029/2007WR006695, 2009.
Wilby, R. L.: Uncertainty in water resource model parameters used for climate change impact assessment, Hydrol. Process., 19, 3201–3219, 2005.
Xu, C. Y.: Operational testing of a water balance model for predicting climate change impacts, Agr. Forest Meteorol., 98, 295–304, 1999.
Zhang, L., Dawes, W. R., and Walker, G. R.: Response of mean annual evapotranspiration to vegetation changes at catchment scale, Water Resour. Res., 37, 701–708, 2001.
Zhang, L., Hickel, K., Dawes, W. R., Chiew, F. H. S., Western, A. W., and Briggs, P. R.: A rational function approach for estimating mean annual evapotranspiration, Water Resour. Res., 40, W02502, doi:10.1029/2003WR002710, 2004.
Zhang, L., Potter, N., Zhang, Y. Q., Hickel, K., and Shao, Q. X.: Water balance modeling over variable time scales based on the Budyko framework: model development and testing, J. Hydrol., 360, 117–131, 2008.

HESSD
8, 8701–8736, 2011

Zhang, Y. Q., Chiew, F. H. S., Zhang, L., Leuning, R., and Cleugh, H. A.: Estimating catchment evaporation and runoff using MODIS leaf area index and the Penman-Monteith equation, Water Resour. Res., 44, W10420, doi:10.1029/2007WR006563, 2008.

Zhang, Y. Q., Chiew, F. H. S., Zhang, L., and Li, H. X.: Use of remotely sensed actual evapotranspiration to improve rainfall-runoff modelling in southeast Australia, J. Hydrometeorol., 10, 969–980, doi:10.1175/2009JHM1061.1., 2009.

Table 1. Ranges of parameter values in DWBM (– indicates dimensionless).

Parameter	Units	Description	Lower bound	Upper bound
α_1	–	retention efficiency	1	5
α_2	–	evapotranspiration efficiency	1	5
S_{max}	mm	soil water storage capacity	5	500
d	–	baseflow linear regression	0.01	1

HESSD
8, 8701–8736, 2011

The transferability of hydrological models under nonstationary climatic conditions

C. Z. Li et al.

HESSD
8, 8701–8736, 2011

Table 2. Ranges of parameter values in the SIMHYD model (– indicates dimensionless).

Parameter	Units	Description	Lower bound	Upper bound
INSC	mm	interception store capacity	0.5	5.0
COEFF	mm	maximum infiltration loss	50	400
SQ	–	infiltration loss exponent	0	6.0
SMSC	mm	soil moisture store capacity	50	500
SUB	–	constant of proportionality in interflow equation	0	1
CRAK	–	constant of proportionality in groundwater recharge equation	0	1
K	–	baseflow linear regression parameter	0.003	0.3

Table 3. Calibration results for the two rainfall-runoff models under different calibration periods.

Indicator	SIMHYD calibrated on dry period	SIMHYD calibrated on wet period	DWBM calibrated on dry period	DWBM calibrated on wet period
25th NSE	0.84	0.85	0.71	0.77
Median NSE	0.70	0.77	0.58	0.66
75th NSE	0.61	0.68	0.43	0.54
Average NSE	0.70	0.76	0.57	0.65
25th WBE	22	16	25	24
Median WBE	13	8	15	12
75th WBE	6	4	9	5
Average WBE	14	11	22	17

HESSD
8, 8701–8736, 2011

The transferability of hydrological models under nonstationary climatic conditions

C. Z. Li et al.

Table 4. Results of model validation under different calibration periods.

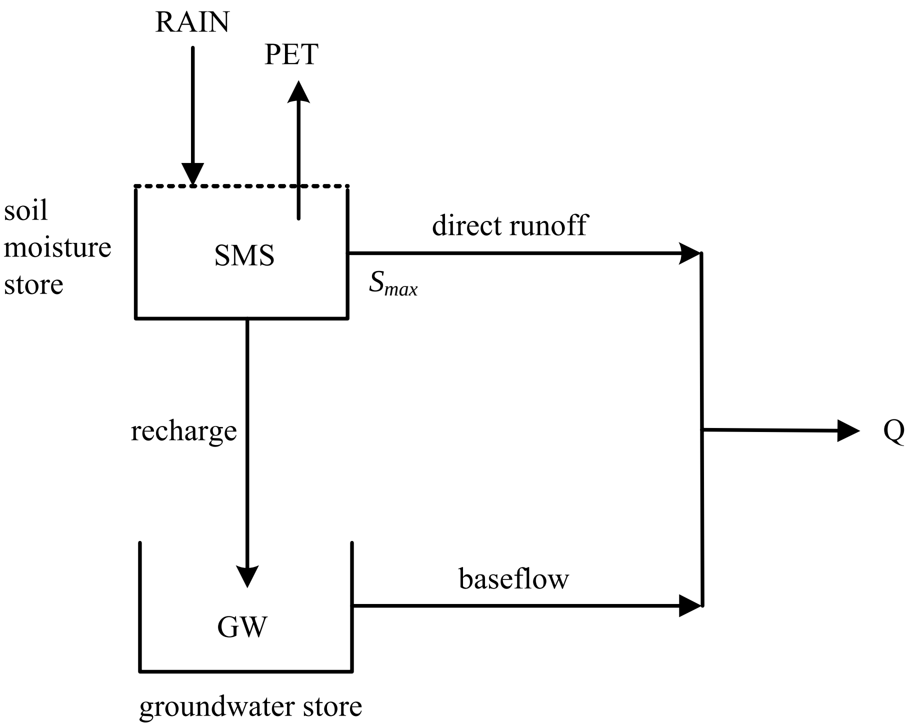
Model	Indicator	dry/dry	dry/wet	wet/dry	wet/wet
SIMHYD	25th NSE	0.72	0.74	0.68	0.77
	Median NSE	0.55	0.64	0.51	0.69
	75th NSE	0.42	0.44	0.41	0.55
	Average NSE	0.57	0.61	0.54	0.66
	25th WBE	34	30	39	23
	Median WBE	20	19	28	13
	75th WBE	14	8	16	7
	Average WBE	24	21	29	17
DWBM	25th NSE	0.56	0.65	0.51	0.72
	Median NSE	0.46	0.48	0.45	0.61
	75th NSE	0.34	0.35	0.30	0.42
	Average NSE	0.48	0.52	0.45	0.59
	25th WBE	35	29	53	25
	Median WBE	22	20	33	18
	75th WBE	15	12	18	11
	Average WBE	27	23	36	19

HESSD
8, 8701–8736, 2011

Table 5. Percent of the catchments in which the model parameter distributions for a dry and wet calibration period were significantly different ($p < 0.01$) under Monte Carlo simulation. Also shown are the results for water-limited ($E_p/P > 1$) and energy-limited ($E_p/P < 1$) catchments. For each model, the parameters are ranked from the most sensitive to calibration conditions to least sensitive.

Model	Parameter	Percent of catchments	Percent of water-limited catchments	Percent of energy-limited catchments
SIMHYD	SUB	63	81	43
	SMSC	60	75	43
	SQ	53	56	50
	CRAK	50	63	36
	K	37	31	43
	COEFF	33	38	29
	INSC	10	13	7
DWBM	α_1	67	81	50
	S_{max}	63	75	50
	d	47	63	29
	α_2	23	25	21

HESSD
8, 8701–8736, 2011

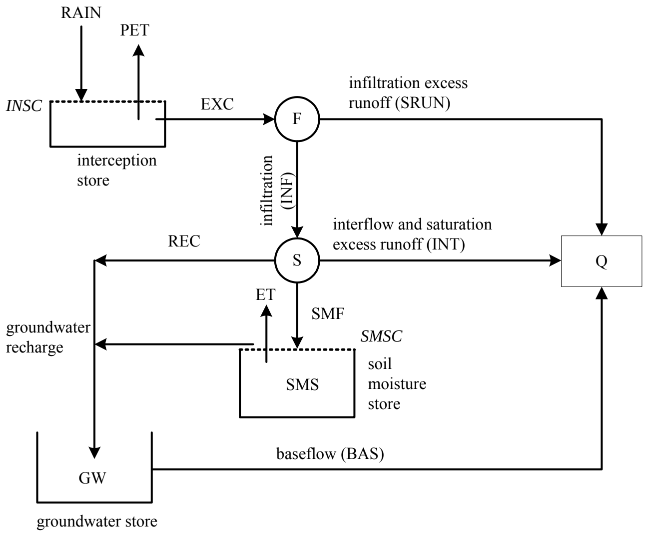


Model parameters and description

α_1	retention efficiency
α_2	evapotranspiration efficiency
S_{max}	soil water storage capacity (mm)
d	baseflow linear regression

Fig. 1. Structure of the lumped water balance model DWBM.

HESSD
8, 8701–8736, 2011



PET = areal potential evapotranspiration (input data)
EXC = RAIN - *INSC*, EXC > 0
INF = lesser of { COEFF exp (-*SQ*×SMS/*SMSC*) , EXC }
SRUN = EXC - INF
INT = *SUB* × SMS/*SMSC* × INF
REC = *CRAK* × SMS/*SMSC* × (INF - INT)
SMF = INF - INT - REC
ET = lesser of { 10 × SMS/*SMSC* , PET }
BAS = *K* × GW

Model parameters and description

INSC	interception store capacity (mm)
COEFF	maximum infiltration loss (mm)
SQ	infiltration loss exponent
SMSC	soil moisture store capacity (mm)
SUB	constant of proportionality in interflow equation
CRAK	constant of proportionality in groundwater recharge equation
K	baseflow linear recession parameter

Fig. 2. Structure of the lumped daily rainfall-runoff model SIMHYD.

HESSD
8, 8701–8736, 2011

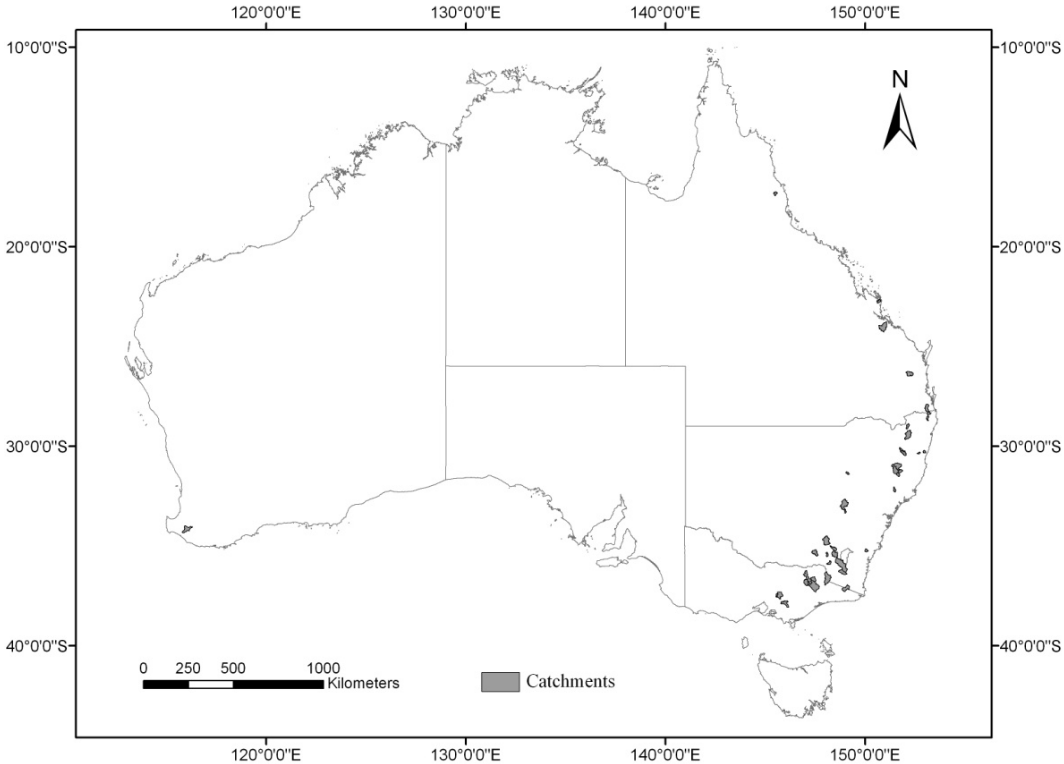


Fig. 3. Location of the 30 catchments used for this study.

HESSD

8, 8701–8736, 2011

The transferability of hydrological models under nonstationary climatic conditions

C. Z. Li et al.

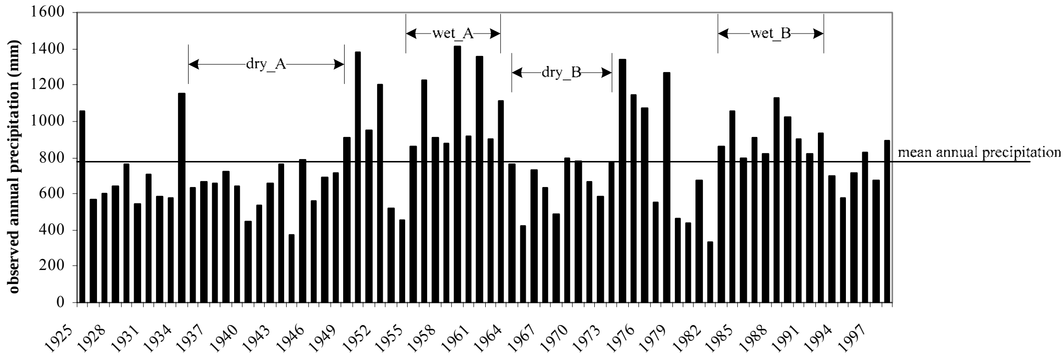


Fig. 4. How the historical rainfall record was divided into 2 wet phases **(A)** and 2 dry phases **(B)** to represent different calibration conditions for the Corang River catchment.

HESSD

8, 8701–8736, 2011

The transferability of hydrological models under nonstationary climatic conditions

C. Z. Li et al.

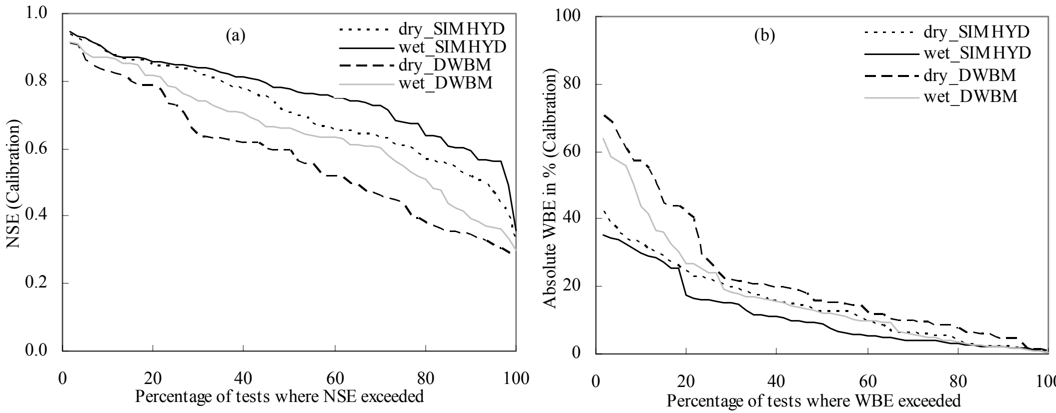


Fig. 5. Summary of calibrated Nash-Sutcliffe Efficiency (NSE) and absolute percentage Water Balance Error (WBE) values for the two rainfall-runoff models under different calibration periods.

HESSD
8, 8701–8736, 2011

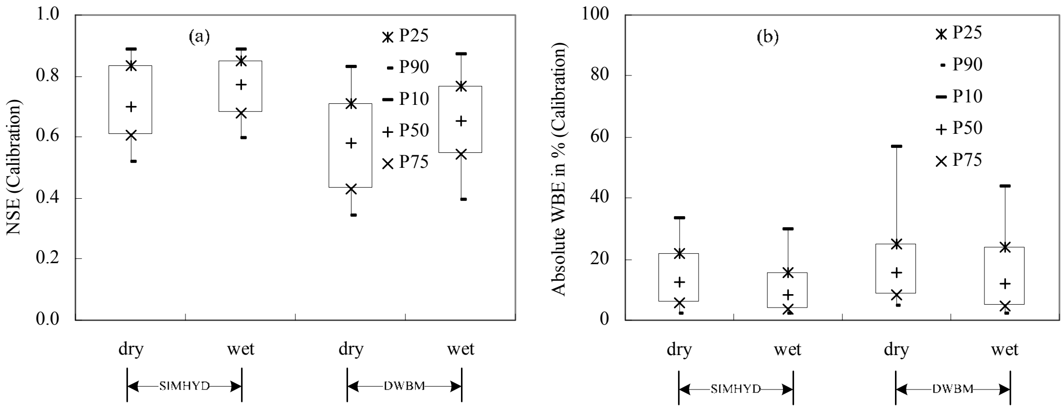


Fig. 6. Summary of Nash-Sutcliffe Efficiency (NSE) and absolute percentage Water Balance Error (WBE) values for the two rainfall-runoff models under different calibration periods. Box-whisker plots show 10th, 25th, 50th (median), 75th, and 90th percentile of results.

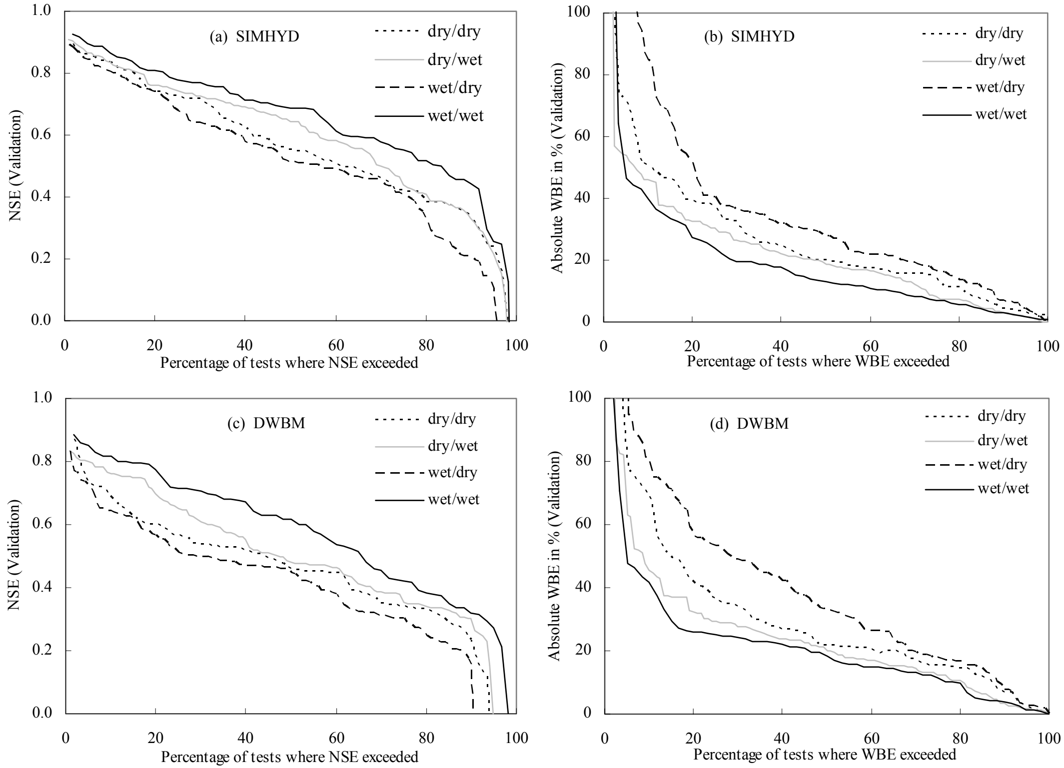


Fig. 7. Summary of validated Nash-Sutcliffe Efficiency (NSE) and absolute percentage Water Balance Error (WBE) values for the two rainfall-runoff models under different calibration periods.

HESSD
8, 8701–8736, 2011

The transferability of hydrological models under nonstationary climatic conditions

C. Z. Li et al.

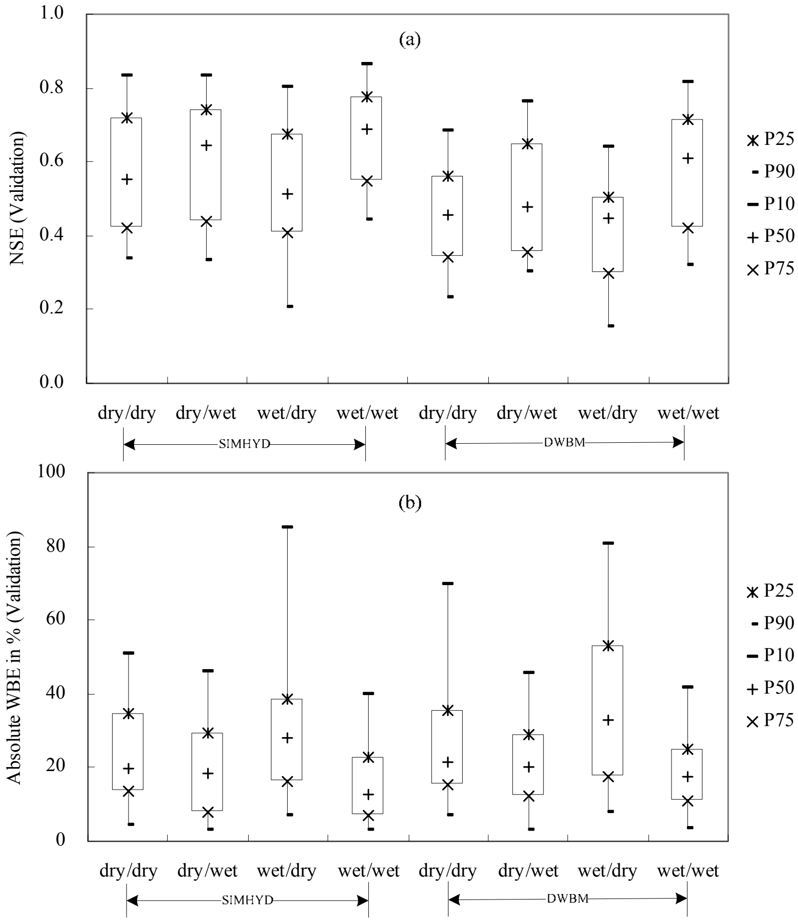


Fig. 8. Summary of validated Nash-Sutcliffe Efficiency (NSE) and absolute percentage Water Balance Error (WBE) values for the two rainfall-runoff models under different calibration periods. Box-whisker plots show 10th, 25th, 50th (median), 75th, and 90th percentile of results.

HESSD
8, 8701–8736, 2011