AGU PUBLICATIONS

Geophysical Research Letters

RESEARCH LETTER

10.1002/2015GL067326

Key Points:
- Hydrological model performance often degrades during prolonged shifts in climate
- Climate shifts sometimes lead to changes in internal catchment functioning
- Models perform poorly and become strongly biased where such changes occur, but not otherwise

Supporting Information:
- Supporting Information S1

Correspondence to:
M. Saft,
margarita.saft@unimelb.edu.au

Citation:
Saft, M., M. C. Peel, A. W. Western, J.-M. Perraud, and L. Zhang (2016), Bias in streamflow projections due to climate-induced shifts in catchment response, *Geophys. Res. Lett.*, *43*, 1574–1581, doi:10.1002/2015GL067326.

Received 22 DEC 2015
Accepted 25 JAN 2016
Accepted article online 26 JAN 2016
Published online 19 FEB 2016

Bias in streamflow projections due to climate-induced shifts in catchment response

Margarita Saft[1,2], Murray C. Peel[1], Andrew W. Western[1], Jean-Michel Perraud[2], and Lu Zhang[2]

[1]Department of Infrastructure Engineering, University of Melbourne, Melbourne, Victoria, Australia, [2]CSIRO Land and Water, Canberra, ACT, Australia

Abstract Demand for quantitative assessments of likely climate change impact on runoff is increasing and conceptual rainfall-runoff models are essential tools for this task. However, the capacity of these models to extrapolate under changing climatic conditions is questionable. A number of studies have found that model predictive skill decreases with changed climatic conditions, especially when predicting drier climates. We found that model skill only declines under certain circumstances, in particular, when a catchment's rainfall-runoff processes change due to changed climatic drivers. In catchments where the rainfall-runoff relationship changed significantly in response to prolonged dry conditions, runoff was consistently overestimated. In contrast, modeled runoff was unbiased in catchments where the rainfall-runoff relationship remained unchanged during the dry period. These conclusions were not model dependent. Our results suggest that current projections of runoff under climate change may provide overly optimistic assessments of future water availability in some regions expecting rainfall reductions.

1. Introduction

Conceptual rainfall-runoff models are routinely used for climate change impact assessments on water resources. These models use meteorological data, typically rainfall and potential evapotranspiration, to predict catchment runoff and are calibrated against observed streamflow records. An assumption of climate change impact assessments is that conceptual rainfall-runoff models perform reasonably well under modified climate conditions. *Klemeš* [1986] was first to highlight the need to test this assumption and introduced the idea of evaluating model performance not only on an independent period but also on a period of different climatic conditions to the calibration period. Now almost 30 years later, evaluating hydrological model performance on an independent period is common in hydrological climate change studies, whereas evaluation in a contrasting climate is not. Diagnostic studies exploring the transferability of model parameters in time and to periods of different climate unanimously conclude that hydrological model parameters are sensitive to the climatic conditions of the calibration period [*Chiew et al.*, 2009; *Chiew et al.*, 2014; *Coron et al.*, 2012; *Li et al.*, 2012; *Merz et al.*, 2011; *Seibert*, 2003; *Vaze et al.*, 2010; *Wagener et al.*, 2003; *Wilby*, 2005]. For example, *Merz et al.* [2011] observed temporal trends in model parameters and related them to trends in evapotranspiration and prevailing catchment wetness conditions, while *Osuch et al.* [2015] found that model parameters were correlated to several climatic indices. Other studies found that model performance degradation was directly related to the difference in precipitation between calibration and evaluation periods [*Coron et al.*, 2012; *Vaze et al.*, 2010]. However, the magnitude of performance loss varied greatly between catchments and the scatter in results remains unexplained.

One plausible explanation for the wide range of model performance loss between catchments could be differences in catchment hydrological response to changed conditions. Modification of catchment processes is acknowledged as possible, or even likely, under altered climates [*Blöschl and Montanari*, 2010; *Chiew et al.*, 2014; *Coron et al.*, 2012; *Peel and Blöschl*, 2011; *Wagener et al.*, 2010] and the likelihood of such change may depend on the biophysical properties of a catchment. One way to explore this issue is to utilize large-scale nature-driven experiments [*Wagener et al.*, 2010], such as the recent prolonged dry period in southeastern Australia (circa 1997–2008) [*Potter and Chiew*, 2011]. In other work [*Saft et al.*, 2015], we investigated changes in hydrologic response in a comprehensive set of largely unimpaired Australian catchments, focusing on pinpointing internal (to the catchment) change in rainfall-runoff processes associated with observed climatic shifts. *Saft et al.* [2015] empirically demonstrated that climatic shifts such as the recent Millennium drought can alter internal catchment functioning causing further runoff reductions in addition to direct impacts of changed precipitation and potential evapotranspiration. This finding supports the idea of long-term catchment adaptation to climate, which is theoretically well grounded in systems theory (catchment coevolution paradigm [*Troch et al.*, 2015]) but rarely demonstrated through historical hydrological observations.

This study focuses on the question of whether model performance degradation is purely due to different climatic forcing or to changed catchment functioning in response to different climatic forcing. We utilize the recent prolonged dry period in southeastern Australia (circa 1997–2008) as an observed example for exploring model robustness under contrasting climate conditions that may also be representative of future conditions. The experimental setup is similar to approaches commonly used in modeling climate change impact in order to gain new insights into how confident we can be in the hydrological component of climate change impact assessment studies. We demonstrate that a catchment's susceptibility to change should be considered when making projections for climatically different future conditions.

2. Methods

We tested conceptual rainfall-runoff models of varying complexity and structure by calibrating them on all available non-Millennium drought data and evaluating their predictive performance over the Millennium drought. The Millennium drought included sustained dry periods of 7 years or longer (based on the rainfall record) in all catchments used. These results were analyzed in conjunction with results of statistical tests of observed changes in the rainfall-runoff relationships [*Saft et al.*, 2015]. For more details on the data, catchment set, dry period definition, and statistical testing, please refer to the supporting information and *Saft et al*. [2015]. Sections 2.2 and 2.3 below provide a brief summary.

2.1. Modeling Setup

We employed six lumped conceptual rainfall-runoff models (Sacramento [*Burnash et al*., 1973], GR4J [*Perrin et al*., 2003], SIMHYD [*Chiew et al*., 2002], SMARG [*Goswami et al*., 2002], AWBM [*Boughton*, 2004], and IHACRES [*Croke et al*., 2006]), which are widely used in Australia and across the world. Conceptual rainfall-runoff models were chosen, as this model type is typical for both operational water management and climate change impact assessments. The models range in level of complexity (e.g., 4–18 parameters) and structure.

The calibration-evaluation approach adopted here is a variant of the split sample test. Split sample testing is a technique used to investigate the transferability of parameters from one period to another [*Klemeš*, 1986]. In our case all available record, except the dry period of interest, is used for the calibration. Hence, the calibration period is typically several times longer than the dry period used for the evaluation, and it also usually includes a number of dry periods of varying length and magnitude. The models were run on a daily time step, and the objective functions and simulation metrics were also calculated on daily data. Model parameters were optimized against a bias-constrained Nash-Sutcliffe Efficiency (NSE) function [*Nash and Sutcliffe*, 1970; *Viney et al*., 2009]. NSE is based on the sum of squared errors and is a commonly used measure of the goodness of fit of hydrological models. The maximum possible value of NSE is 1 when there is no difference between modeled and observed data. Negative NSE indicates that modeled results are worse than using the average value of the time series. The bias constraint is to ensure that the optimal parameter set, identified through calibration, leads to low systematic error during the calibration period. Bias was defined as the difference between the modeled and observed streamflow divided by the observed streamflow. The model evaluation assessment statistics (NSE and relative bias) were calculated for the dry period of interest, illustrating the model predictive performance and parameter adequacy.

2.2. Shift in the Rainfall-Runoff Relationship

The modeling results were analyzed in conjunction with results obtained from statistical tests of whether the annual rainfall-runoff relationship shifted significantly during the dry period of interest compared with the historical record excluding the dry period of interest (Millennium drought) [*Saft et al*., 2015]. The annual rainfall-runoff relationship is the functional dependence of annual runoff on annual rainfall. It captures catchment response during wet, moderate, or dry years. The typical catchment response experienced during dry years forms the lower part of this curve. To test for a shift in annual rainfall-runoff relationship, a *t* test was applied to the dry period indicator variable in a linear regression relationship between rainfall and Box-Cox

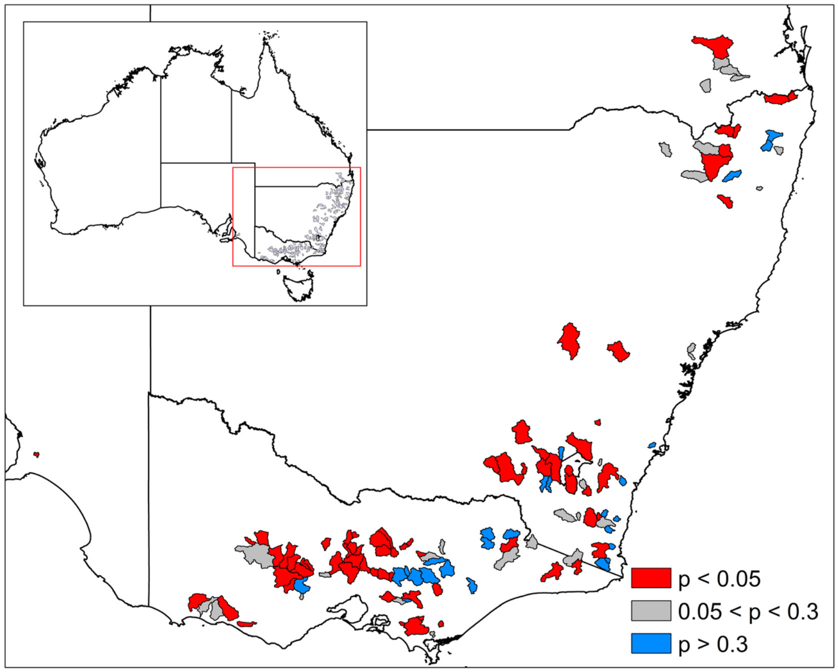

Figure 1. Location of the study catchments with statistical test results for shifts in the rainfall-runoff relationships during the Millennium drought ($p < 0.05$ = "shift" and $p > 0.3$ = "no shift").

transformed runoff (more details are provided in the supporting information). The rainfall-runoff relationship does not explicitly include the influence of temperature or potential evapotranspiration, although these factors contribute to the scatter around the fitted relationship. Nevertheless, the interaction between potential evapotranspiration and rainfall is already captured by the historical annual rainfall-runoff relationship; because it is an empirical relationship. Furthermore, potential evapotranspiration (Morton's areal PET) anomalies during the Millennium drought for our catchment set were small (~1–2% on average) and not statistically different between catchments with and without change in the rainfall-runoff relationship [*Saft et al.*, 2015]. Several other dry period characteristics including drought length, dry period rainfall anomaly, wet day rainfall anomaly, percent change in seasonality, and in monthly and annual rainfall C_v were compared between catchments with and without a shift in the rainfall-runoff relationship. None of these characteristics were found to be significantly different between the two groups of catchments. Therefore, neither changes in rainfall patterns nor the severity of the drought in particular catchments could explain the shifts; whereas significant differences were detected between biophysical properties of the two groups of catchments (e.g., slope, percent of woody cover, and historical climate aridity). Based on this we consider a shift in the annual rainfall-runoff relationship to be an indicator of change in internal catchment processes occurring over the extended dry period.

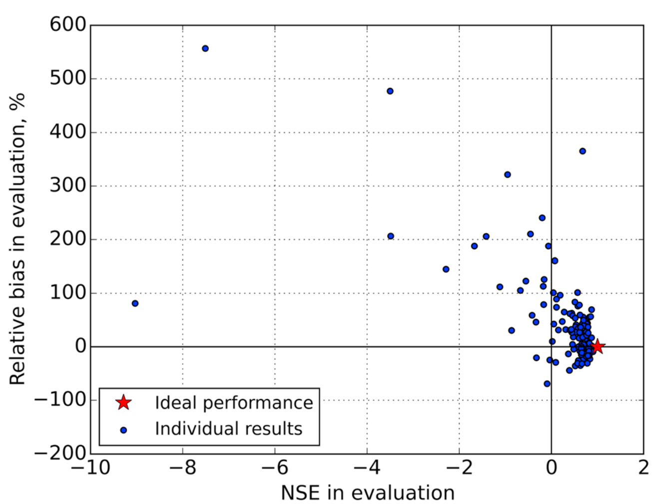

Figure 2. Example of scatter in model predictive performance for a set of catchments: Sacramento Model.

2.3. Data Set

Figure 1 shows the catchments considered in this study. They are largely unimpaired catchments with no major artificial influences on the flow, such

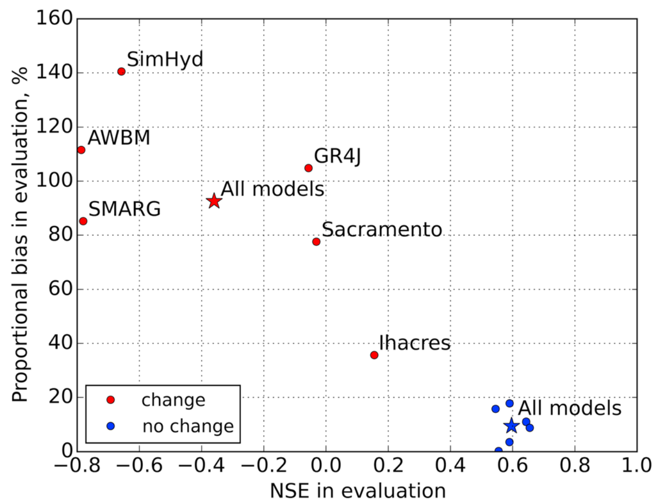

Figure 3. Average evaluation results (NSE versus proportional bias) during the dry period for catchments with and without a change in rainfall-runoff relationship.

as reservoirs or irrigation schemes. We used the 124 catchments from *Saft et al*. [2015] in which the Millennium drought was observed without interruption (i.e., dry period length ≥7 years). Figure 1 shows three groups of catchments based on the significance testing outlined above. The no-shift group has catchments with $p > 0.3$, the intermediate group has $0.3 > p > 0.05$, and the significant shift group has $p < 0.05$. When we compare catchments clearly with a shift and clearly without a shift we compare the two more extreme groups. We considered that the existence or otherwise of a shift for the catchments in the intermediate group was unclear, and hence, we excluded them to improve the robustness of the comparison.

3. Results

Here we demonstrate that decline in hydrologic model predictive performance is associated with changes in rainfall-runoff processes in response to the extended dry period, rather than just the direct reduction in rainfall. Consistent with previous studies [*Coron et al*., 2012; *Vaze et al*., 2010], our results exhibit a wide range of model performance (see Figure 2 for Sacramento model). The performance is considered best when the relative bias is close to zero (no overprediction or underprediction) and when the Nash-Sutcliffe Efficiency (NSE) approaches 1. To understand the origin of the wide range of model predictive performance during the dry period (Figure 2), we separately analyzed two groups of catchments: one with a statistically significant shift in rainfall-runoff relationship during the dry period and the other without.

Our results suggest that despite the severe and prolonged dry period, all models performed well in catchments where no change in rainfall-runoff relationship was observed. Figure 3 shows that the average evaluation NSE values calculated from daily data are close to 0.6 for all models, and relative bias is between zero and 20%

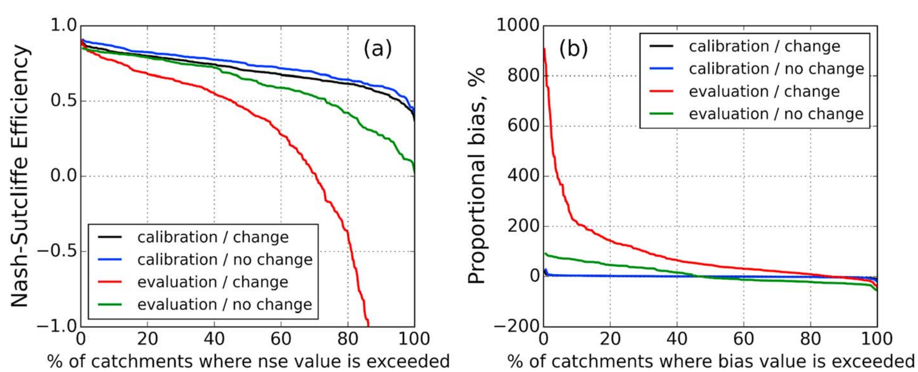

Figure 4. Exceedance curves for the (a) NSE and (b) proportional bias for all models and catchments combined.

(depending on the model), with an average bias across all models below 10%. Slight positive bias is common when simulating dry periods (this is mostly explained by the fact that common objective functions which use sum of squared errors such as NSE put more weight on high and moderate flows). Model performance during evaluation is expected to be lower than that during calibration, which is shown by the difference between the blue line and the green line in Figure 4a for catchments without a shift in the rainfall-runoff relationship. A decline in model performance of this magnitude is common for split sample tests of an independent period. Further evidence of good model performance for catchments without a shift in rainfall-runoff relationship is that bias has a normal distribution centered on zero with symmetrical and short tails (Figure 4b, green line).

The results for catchments where the prolonged dry period led to a shift in the annual rainfall-runoff relationship are in stark contrast to those discussed above. Where the rainfall-runoff relationship shifted, model performance was poor (Figures 3 and 4). The average evaluation NSE for all models with change is negative, and the results for the best performing models (Sacramento and IHACRES) are below 0.2. Bias is positive in over 80% of these catchments (Figure 4b), indicating that the models tend to overestimate flow during the dry period. Average bias for the ensemble of models is 93% (Figure 3), which means that flow was overestimated; being almost double the observed, on average. Individual model results (provided in the supporting information) show similar patterns. Thus, we cannot attribute the exacerbated decline in model performance to a deficiency in a particular model. Rather, this issue appears to be a systematic problem of conceptual rainfall-runoff models in general.

4. Discussion

The contrast in results from the two groups of catchments indicates that reduced model predictive performance during a long dry period depends on whether or not a shift in the rainfall-runoff response occurs. In about half of the cases the catchment response to short-term dry periods (up to a few years) was not representative of the long-term response (7 years or longer), and hence, short dry periods did not properly inform the model fitting. Thus, reduction in model performance was caused by the change in internal catchment dynamics during multiyear dry sequences, and not by the decrease in rainfall itself, nor by changes in temperature and/or potential evapotranspiration.

The performance degradation was fundamentally about bias and not just noise. Similarity in model behavior (e.g., consistent direction of bias between models) means that the models might provide relatively similar yet largely inaccurate predictions. For example, in cases where the rainfall-runoff relationship shifted, the six models produced flow estimates with bias between ~ +36% and ~ +140%, with an average bias of ~ +93%. Without knowing the true bias, the variation from the model ensemble average would appear to be ±25–30% of the predicted average, which is very misleading. Put another way, for the great majority of catchments with change, the observation actually fell below the range of the six models. Therefore, even an ensemble of models (one of the suggested options to add more confidence in modeling results [*Seiller et al.*, 2012]) overestimates the flow by over 90% on average and does not provide a reliable uncertainty estimate.

Our study results have implications for alternative ways of calibrating models to better deal with changing climate. One way would be to calibrate against as long a record as possible with both wet and dry periods to train the model across a variety of climatic conditions. Our study demonstrates that even long records with a number of dry spells may not be sufficient to train a model to deal with a persistent change in precipitation. Another approach is to select a calibration period based on the expected future climate, e.g., to calibrate against historical dry periods if the climate is expected to become drier [*Li et al.*, 2012]. But, if short dry periods do not prompt a change in rainfall-runoff relationship and hydrological behavior, then calibrating rainfall-runoff models on short dry periods should not be assumed to be sufficient for use in climate change studies without further investigation of potential changes in hydrological processes within the catchment.

Our results also indicate that uncertainty from hydrologic modeling is likely to be underestimated when a change in rainfall-runoff relationship is prompted by a projected climate change. Uncertainty arising from different parts of hydrological climate change impact assessment (emission scenario, global climate model, downscaling method, natural climate variability, and hydrological model) is often judged by the spread of outcomes arising from each stage, utilizing essentially sensitivity-like methods. With this general approach, numerous studies have suggested that Global Climate Models dominate the uncertainty, sometimes with significant contributions from downscaling methods, whereas hydrological modeling contributes little uncertainty [*Arnell*, 2011; *Chiew et al.*, 2009; *Kay et al*., 2009; *Kingston and Taylor*, 2010; *Prudhomme and Davies*, 2009; *Teng et al*., 2012; *Wilby and Harris*, 2006]. However, in recent years studies have reported that uncertainty originating from hydrological models can be substantial (in some cases it is comparable to major components of the climate modeling) and therefore should not be overlooked [*Bosshard et al*., 2013; *Dams et al*., 2015; *Honti et al*., 2014; *Jung et al*., 2012; *Lespinas et al*., 2014; *Velázquez et al*., 2013]. Our results add to the latter view by stressing that precision in model outputs does not necessarily indicate low uncertainty in the cases of systematic bias.

From the perspective of climate change impact assessment studies, our results are concerning as they indicate likely overestimation of available water resources in the future in some regions. Many climate projections indicate increased temperature, decreased rainfall, and more frequent droughts in midlatitude and subtropical arid and semiarid areas around the world [*Arnell*, 1999; *Hewitson et al*., 2014; *Intergovernmental Panel on Climate Change*, 2013; *Tang and Lettenmaier*, 2012]. Areas potentially affected include southern Europe, the Middle East, southern Africa, the Southwest United States, and southeastern and southwestern Australia [*Arnell*, 1999; *Milly et al*., 2005]. Based on our assessment of catchments in southeastern Australia where shifts in the rainfall-runoff relationship were associated with drier catchments [*Saft et al*., 2015], it is likely that these issues will appear first in already water-scarce areas. On the other hand, wetter catchments were less likely to experience the change in hydrologic response that resulted in degradation of model predictive performance. Thus, the main runoff generating areas may be more accurately modeled even during climate variations.

To improve hydrologic modeling of prolonged dry periods, a better understanding of the physical mechanisms driving shifts in catchment response is required. While detailed process interpretation is beyond the scope of this study, we would like to note that several physical mechanisms relating to groundwater, vegetation, and soil properties have been suggested to influence the rainfall-runoff response in the literature. Examples include increased runoff coefficients in the Sahelian region resulting from vegetation degradation and soil crusting during multidecadal drought [*Descroix et al*., 2009] and groundwater decline amplifying streamflow reduction in southwestern Australia [*Hughes et al*., 2012; *Petrone et al*., 2010]. *Petheram et al*. [2011] and *Chiew et al*. [2014] have postulated that interrupted connection between shallow groundwater and soil moisture exacerbated streamflow declines during the Millennium drought; however, no detailed studies have been undertaken, so the cause remains unclear. It is likely that the responsible mechanisms may vary between catchments and may involve various groundwater, vegetation, soil, and/or other effects.

This study has been based on prolonged but temporary shifts in climate that have been observed, and which resulted in a shift in the rainfall-runoff relationship in 56.5% of the catchments analyzed [*Saft et al*., 2015]. In the future we might experience permanent shifts, as well as longer more severe dry sequences, which may induce larger or more widespread modifications of rainfall-runoff processes that also impact more resilient catchments.

5. Conclusion

Typical conceptual rainfall-runoff models are able to predict runoff accurately even under changed climate in some of the study catchments but are heavily biased in other catchments. Models perform well in catchments where the hydrologic response during a decade-long dry period was similar to shorter droughts. However, when the response to prolonged dry period differs from response to similar magnitude but shorter duration variations (i.e., isolated dry years or shorter sequences of dry years) in external forcing, then output of conventional hydrological models contains systematic bias. If a projected climate change would induce a downward shift in the rainfall-runoff relationship, then streamflow projections using current methods are shown to provide overly optimistic assessments of water availability during a time of declining water resources, further increasing the challenge for water managers. Potential modification of catchment processes during an extended change in climate needs to be assessed in order to provide more reliable estimates. In view of the above, further research is required to identify what drives changes in rainfall-runoff processes during climatically different periods and how to predict such changes. Improved understanding of how hydrologic processes are modified by changing climatic conditions will be a significant contribution to the International Association of Hydrological Sciences decade Panta Rhei [*Montanari et al*., 2013].

Acknowledgments
The research project was funded by the CSIRO Water for a Healthy Country National Research Flagship. Murray Peel is the recipient of an Australian Research Council Future Fellowship (FT120100130). We thank Jai Vaze for providing streamflow data and Fangfang Zhao for sharing rainfall data. Source rainfall and PET data were extracted from SILO database (https://www.longpaddock.qld.gov.au/silo/). Streamflow data are publicly available from the relevant government state agencies online. Hydrologic models were coded in C# as part of the CSIRO "Catchment Water Yield Estimation Tool" project.

References

Arnell, N. W. (1999), Climate change and global water resources, *Global Environ. Change*, *9*(Supplement 1), S31–S49.

Arnell, N. W. (2011), Uncertainty in the relationship between climate forcing and hydrological response in UK catchments, *Hydrol. Earth Syst. Sci.*, *15*(3), 897–912.

Blöschl, G., and A. Montanari (2010), Climate change impacts—Throwing the dice?, *Hydrol. Processes*, *24*(3), 374–381.

Bosshard, T., M. Carambia, K. Goergen, S. Kotlarski, P. Krahe, M. Zappa, and C. Schär (2013), Quantifying uncertainty sources in an ensemble of hydrological climate-impact projections, *Water Resour. Res.*, *49*, 1523–1536, doi:10.1029/2011WR011533.

Boughton, W. (2004), The Australian water balance model, *Environ. Modell. Software*, *19*(10), 943–956.

Burnash, R. J., R. L. Ferral, and R. A. McGuire (1973), A generalized streamflow simulation system, conceptual modeling for digital computers.

Chiew, F. H. S., M. C. Peel, and A. W. Western (2002), Application and testing of the simple rainfall-runoff model SIMHYD, in *Mathematical Models of Small Watershed Hydrology and Applications*, edited by V. P. Singh and D. Frevert, pp. 335–367, Water Resour. Publ., Boulder, Colo.

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

Chiew, F. H. S., N. J. Potter, J. Vaze, C. Petheram, L. Zhang, J. Teng, and D. A. Post (2014), Observed hydrologic non–stationarity in far south-eastern Australia: Implications for modelling and prediction, *Stochastic Environ. Res. Risk Assess.*, *28*(1), 3–15.

Coron, L., V. Andréassian, C. Perrin, J. Lerat, J. Vaze, M. Bourqui, and F. Hendrickx (2012), Crash testing hydrological models in contrasted climate conditions: An experiment on 216 Australian catchments, *Water Resour. Res.*, *48*, W05552, doi:10.1029/2011WR011721.

Croke, B. F. W., F. Andrews, A. J. Jakeman, S. M. Cuddy, and A. Luddy (2006), IHACRES classic plus: A redesign of the IHACRES rainfall-runoff model, *Environ. Modell. Software*, *21*(3), 426–427.

Dams, J., J. Nossent, T. B. Senbeta, P. Willems, and O. Batelaan (2015), Multi-model approach to assess the impact of climate change on runoff, *J. Hydrol.*, *529*, 1601–1616.

Descroix, L., et al. (2009), Spatio-temporal variability of hydrological regimes around the boundaries between Sahelian and Sudanian areas of West Africa: A synthesis, *J. Hydrol.*, *375*(1–2), 90–102.

Goswami, M., K. O'Connor, and A. Shamseldin (2002), Structures and performances of five rainfall-runoff models for continuous river-flow simulation, paper presented at Proceedings of 1st Biennial Meeting of International Environmental Modeling and Software Society, Lugano, Switzerland.

Hewitson, B. C., A. C. Janetos, T. R. Carter, F. Giorgi, R. G. Jones, W.-T. Kwon, L. O. Mearns, E. L. F. Schipper, and M. van Aalst (2014), Regional context, in *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part B: Regional Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by V. R. Barros et al., pp. 1133–1197, Cambridge Univ. Press, Cambridge, U. K., and New York.

Honti, M., A. Scheidegger, and C. Stamm (2014), The importance of hydrological uncertainty assessment methods in climate change impact studies, *Hydrol. Earth Syst. Sci.*, *18*(8), 3301–3317.

Hughes, J. D., K. C. Petrone, and R. P. Silberstein (2012), Drought, groundwater storage and stream flow decline in southwestern Australia, *Geophys. Res. Lett.*, *39*, L03408, doi:10.1029/2011GL050797.

Intergovernmental Panel on Climate Change (2013), Summary for Policymakers, in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by T. F. Stocker et al., Cambridge Univ. Press, Cambridge, U. K., and New York.

Jung, I. W., H. Moradkhani, and H. Chang (2012), Uncertainty assessment of climate change impacts for hydrologically distinct river basins, *J. Hydrol.*, *466–467*, 73–87.

Kay, A. L., H. N. Davies, V. A. Bell, and R. G. Jones (2009), Comparison of uncertainty sources for climate change impacts: Flood frequency in England, *Clim. Change*, *92*(1–2), 41–63.

Kingston, D. G., and R. G. Taylor (2010), Sources of uncertainty in climate change impacts on river discharge and groundwater in a headwater catchment of the Upper Nile Basin, Uganda, *Hydrol. Earth Syst. Sci.*, *14*(7), 1297–1308.

Klemeš, V. (1986), Operational testing of hydrological simulation models / Vérification, en conditions réelles, des modèles de simulation hydrologique, *Hydrol. Sci. J.*, *31*(1), 13–24.

Lespinas, F., W. Ludwig, and S. Heussner (2014), Hydrological and climatic uncertainties associated with modeling the impact of climate change on water resources of small Mediterranean coastal rivers, *J. Hydrol.*, *511*, 403–422.

Li, C. Z., L. Zhang, H. Wang, Y. Q. Zhang, F. L. Yu, and D. H. Yan (2012), The transferability of hydrological models under nonstationary climatic conditions, *Hydrol. Earth Syst. Sci.*, *16*(4), 1239–1254.

Merz, R., J. Parajka, and G. Blöschl (2011), Time stability of catchment model parameters: Implications for climate impact analyses, *Water Resour. Res.*, *47*, W02531, doi:10.1029/2010WR009505.

Milly, P. C. D., K. A. Dunne, and A. V. Vecchia (2005), Global pattern of trends in streamflow and water availability in a changing climate, *Nature (London)*, *438*(7066), 347–350.

Montanari, A., et al. (2013), "Panta Rhei-Everything Flows": Change in hydrology and society-The IAHS Scientific Decade 2013–2022, *Hydrol. Sci. J.*, *58*(6), 1256–1275.

Nash, J. E., and J. V. Sutcliffe (1970), River flow forecasting through conceptual models part I–A discussion of principles, *J. Hydrol.*, *10*(3), 282–290.

Osuch, M., R. J. Romanowicz, and M. J. Booij (2015), The influence of parametric uncertainty on the relationships between HBV model parameters and climatic characteristics, *Hydrol. Sci. J.*, *60*(7–8), 1299–1316.

Peel, M. C., and G. Blöschl (2011), Hydrological modelling in a changing world, *Prog. Phys. Geogr.*, *35*(2), 249–261.

Perrin, C., C. Michel, and V. Andréassian (2003), Improvement of a parsimonious model for streamflow simulation, *J. Hydrol.*, *279*(1–4), 275–289.

Petheram, C., N. Potter, J. Vaze, F. Chiew, and L. Zhang (2011), Towards better understanding of changes in rainfall-runoff relationships during the recent drought in south-eastern Australia, in *MODSIM 2011—19th International Congress on Modelling and Simulation—Sustaining Our Future: Understanding and Living with Uncertainty*, edited by F. Chan, D. Marinova, and R. S. Anderssen, pp. 3622–3628, Modelling and Simulation Society of Australia and New Zealand, Perth.

Petrone, K. C., J. D. Hughes, T. G. Van Niel, and R. P. Silberstein (2010), Streamflow decline in southwestern Australia, 1950–2008, *Geophys. Res. Lett.*, *37*, L11401, doi:10.1029/2010GL043102.

Potter, N. J., and F. H. S. Chiew (2011), An investigation into changes in climate characteristics causing the recent very low runoff in the southern Murray-Darling Basin using rainfall-runoff models, *Water Resour. Res.*, *47*, W00G10, doi:10.1029/2010WR010333.

Prudhomme, C., and H. Davies (2009), Assessing uncertainties in climate change impact analyses on the river flow regimes in the UK. Part 1: Baseline climate, *Clim. Change*, *93*(1–2), 177–195.

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

Seibert, J. (2003), Reliability of model predictions outside calibration conditions, *Nord. Hydrol.*, *34*(5), 477–492.

Seiller, G., F. Anctil, and C. Perrin (2012), Multimodel evaluation of twenty lumped hydrological models under contrasted climate conditions, *Hydrol. Earth Syst. Sci.*, *16*(4), 1171–1189.

Tang, Q., and D. P. Lettenmaier (2012), 21st century runoff sensitivities of major global river basins, *Geophys. Res. Lett.*, *39*, L06403, doi:10.1029/2011GL050834.

Teng, J., J. Vaze, F. H. S. Chiew, B. Wang, and J. M. Perraud (2012), Estimating the relative uncertainties sourced from GCMs and hydrological models in modeling climate change impact on runoff, *J. Hydrometeorol.*, *13*(1), 122–139.

Troch, P. A., T. Lahmers, A. Meira, R. Mukherjee, J. W. Pedersen, T. Roy, and R. Valdés–Pineda (2015), Catchment coevolution: A useful framework for improving predictions of hydrological change?, *Water Resour. Res.*, *51*, 4903–4922, doi:10.1002/2015WR017032.

Vaze, J., D. A. Post, F. H. S. Chiew, J. M. Perraud, N. R. Viney, and J. Teng (2010), Climate non–stationarity—Validity of calibrated rainfall-runoff models for use in climate change studies, *J. Hydrol.*, *394*(3–4), 447–457.

Velázquez, J. A., J. Schmid, S. Ricard, M. J. Muerth, B. Gauvin St-Denis, M. Minville, D. Chaumont, D. Caya, R. Ludwig, and R. Turcotte (2013), An ensemble approach to assess hydrological models" contribution to uncertainties in the analysis of climate change impact on water resources, *Hydrol. Earth Syst. Sci.*, *17*(2), 565–578.

Viney, N. R., J. Perraud, J. Vaze, F. H. S. Chiew, D. A. Post, and A. Yang (2009), The usefulness of bias constraints in model calibration for regionalisation to ungauged catchments, paper presented at 18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation: Interfacing Modelling and Simulation with Mathematical and Computational Sciences, Proceedings.

Wagener, T., N. McIntyre, M. J. Lees, H. S. Wheater, and H. V. Gupta (2003), Towards reduced uncertainty in conceptual rainfall-runoff modelling: Dynamic identifiability analysis, *Hydrol. Processes*, *17*(2), 455–476.

Wagener, T., M. Sivapalan, P. A. Troch, B. L. McGlynn, C. J. Harman, H. V. Gupta, P. Kumar, P. S. C. Rao, N. B. Basu, and J. S. Wilson (2010), The future of hydrology: An evolving science for a changing world, *Water Resour. Res.*, *46*, W05301, doi:10.1029/2009WR008906.

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

Wilby, R. L., and I. Harris (2006), A framework for assessing uncertainties in climate change impacts: Low-flow scenarios for the River Thames, UK, *Water Resour. Res.*, *42*, W02419, doi:10.1029/2005WR004065.

Geophysical Research Letters

Supporting Information for

Bias in streamflow projections due to climate-induced shifts in catchment response

Margarita Saft[1,2], Murray C. Peel[1], Andrew W.Western[1], Jean-Michel Perraud[2], Lu Zhang[2]

[1]Department of Infrastructure Engineering, The University of Melbourne, Melbourne, Victoria, Australia

[2]Water for a Healthy Country Flagship, CSIRO Land and Water, Canberra, ACT, Australia

Contents of this file

1. Figures S1 to S7
2. Tables S1 to S8
3. Text S1 to S2

Introduction

The Supplementary material contains figures similar to Figure 4 in the main text but presented for each model individually, technical details on the calibration approach and on statistical testing of the rainfall-runoff relationship.
Figures S1 to S6 contain exceedance curves for the Nash-Sutcliffe Efficiency (NSE) and proportional bias (similar to Figure 4 in the main text) for the individual models.
Tables S1 to S6 contain model parameter calibration ranges.
Text S1 gives additional details on the calibration procedure.
Table S7 contains parameters of the optimization algorithm.
Figure S7 illustrates the presence/absence of shift in the rainfall-runoff relationship.
Text S2 describes the statistical test to detect the shift in the rainfall-runoff relationship.
Table S8 provides details on the analysis periods and runoff record details.

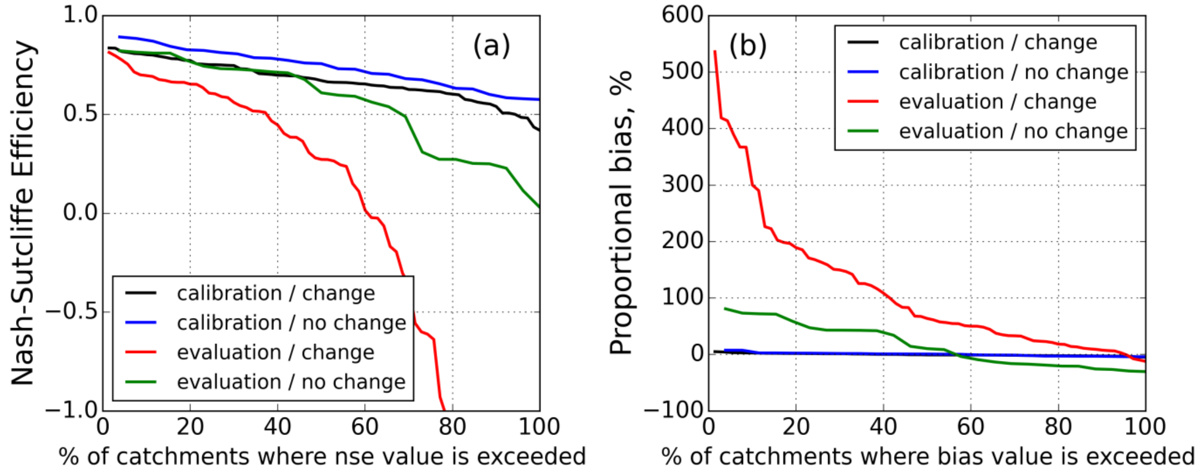

Figure S1. Exceedance curves for the NSE (a) and proportional bias (b) for AWBM model

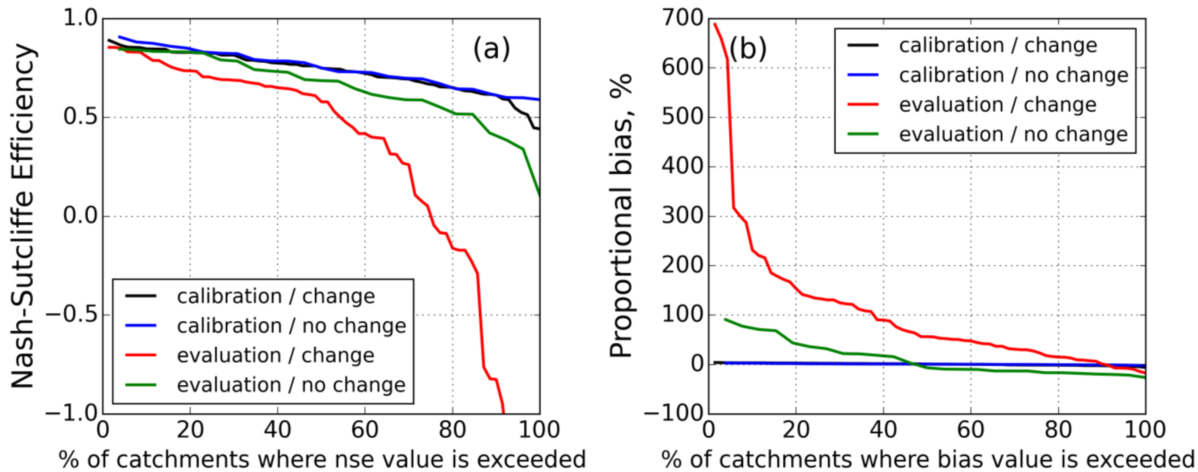

Figure S2. Exceedance curves for the NSE (a) and proportional bias (b) for GR4J model

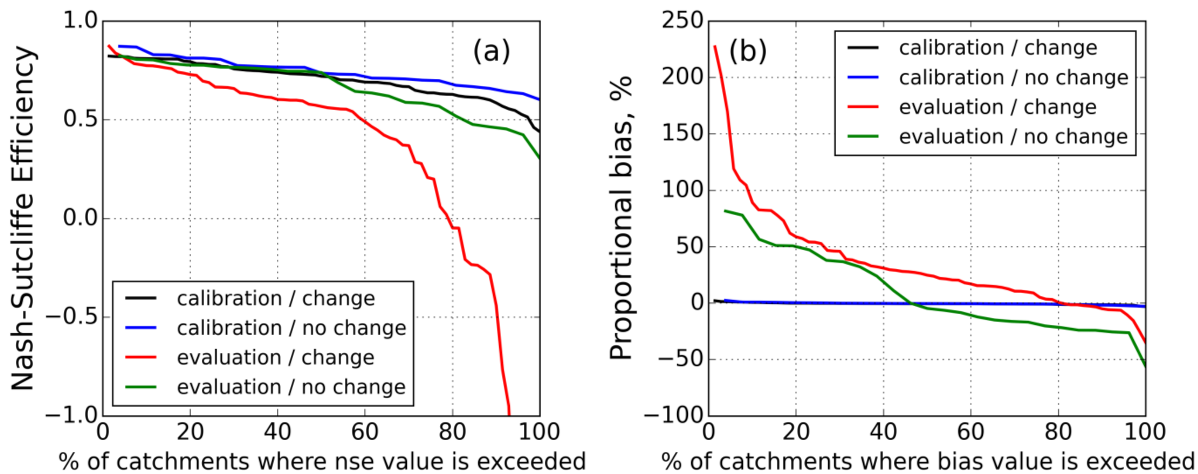

Figure S3. Exceedance curves for the NSE (a) and proportional bias (b) for IHACRES model

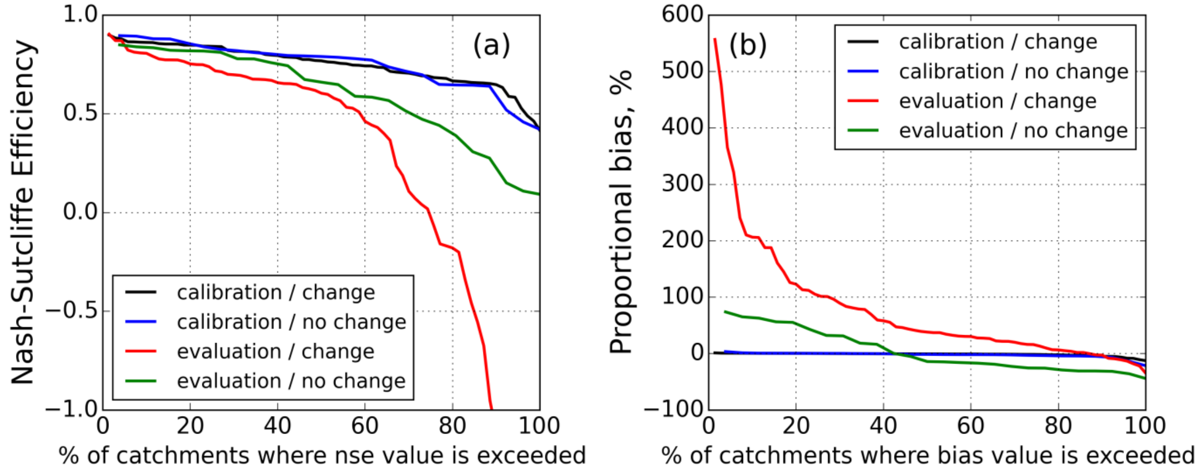

Figure S4. Exceedance curves for the NSE (a) and proportional bias (b) for Sacramento model.

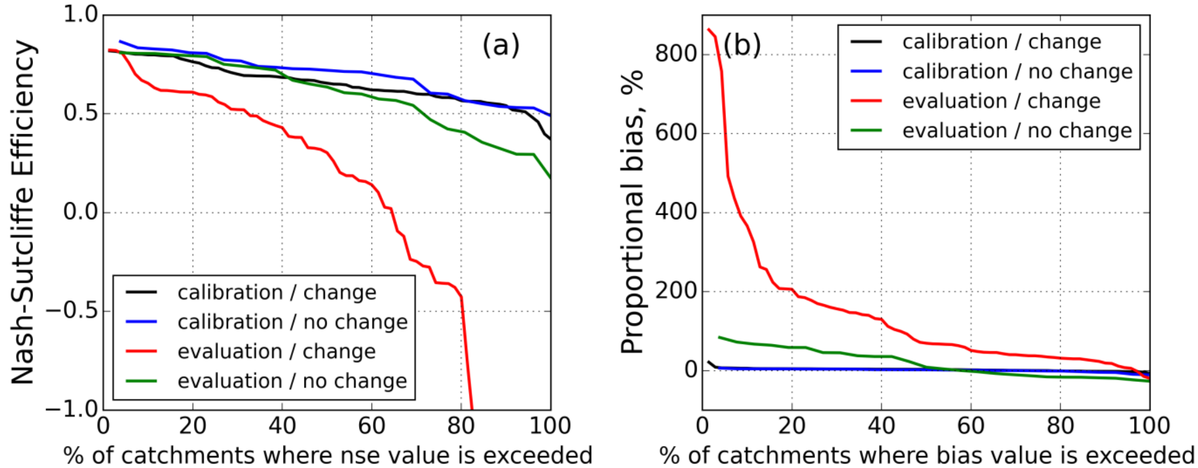

Figure S5. Exceedance curves for the NSE (a) and proportional bias (b) for SIMHYD model

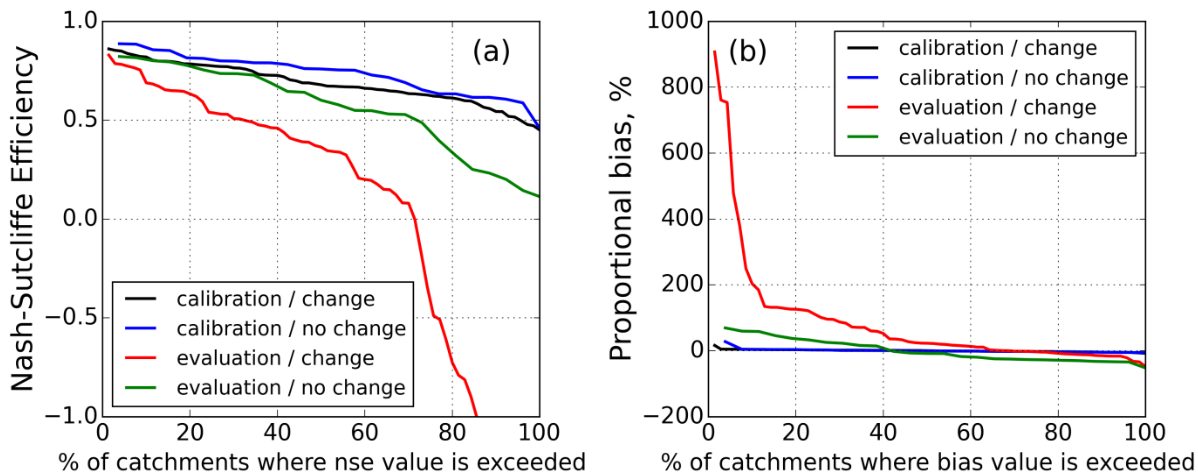

Figure S6. Exceedance curves for the NSE (a) and proportional bias (b) for SMARG model

	Fixed	Calibrated	
Parameter	**Value**	**Min**	**Max**
A1	0.134		
A2	0.433		
BFI		0	1
C1		0	50
C2		0	200
C3		0	500
KBase		0	1
KSurf		0	1

Table S1. Parameter ranges for AWBM model.

Parameter	Min	Max
x1	1	1500
x2	-10	5
x3	1	500
x4	0.5	4

Table S2. Parameter ranges for GR4J model.

	Fixed	Calibrated	
Parameter	**Value**	**Min**	**Max**
F		0	20
InverseC		10	5000
L		0	5000
P	1		
Tq		0	5
Tref	3.4		
Ts		5	500
Tw		1	200
Vs		0	1

Table S3. Parameter ranges for IHACRES model.

	Fixed	Calibrated	
Parameter	**Value**	**Min**	**Max**
Adimp		0	1
Lzfpm		0	100
Lzfsm		0	50
Lzpk		0	0.9
Lzsk		0	1
Lztwm		50	600
Pctim	0		
Pfree		0	0.5
Rexp		1	3
Rserv	0.3		
Sarva		0	0.2
Side	0		
Ssout	0		
Uzfwm		0	100
Uzk		0	1
Uztwm		50	200
Zperc		0	100
UH1		0.01	1
UH2		0	1
UH3		0	1
UH4		0	1
UH5		0	1

Table S4. Parameter ranges for Sacramento model.

	Fixed	Calibrated	
Parameter	**Value**	**Min**	**Max**
Baseflow Coefficient		0	1
Impervious Threshold	0		
Infiltration Coefficient	150		
Infiltration shape	2		
Interflow Coefficient		0	1
K		0.5	200
Pervious Fraction	1		
RISC		0	5
Recharge Coefficient		0	1
SMSC		1	500
X	0		

Table S5. Parameter ranges for SIMHYD model.

	Fixed	Calibrated	
Parameter	**Value**	**Min**	**Max**
Evaporation Coefficient (C)		0	1
Groundwater Runoff Coefficient (G)		0	1
Direct Runoff Coefficient (H)		0	1
Time lag parameter for groundwater Routing (Kg)		1	1000
Number of Linear Reservoirs (N)		1	10
Time lag parameter in Nash Cascade Model (NK)		0.01	10
Potential Evaporation Factor (T)	1		
Soil Moisture Infiltration Rate (Y)		10	200
Soil Moisture Storage Capacity (Z)		0	500

Table S6. Parameter ranges for SMARG model.

Text S1. Details of calibration approach.

The warm-up period was set to 2 years prior to the first available runoff observation. As climatic records were complete, models were run on a daily time step without interruption.

The model parameters were optimized using a global optimization method, Shuffled Complex Evolution (SCE) [*Duan et.al.,* 1994], using settings following the guidelines in that paper. Given the number n of calibrated model parameters, the SCE parameters are shown in Table S7. The convergence criterion was either reaching a maximum coefficient of variation in the SCE population of less than 2.5% for all parameters, or 10 hours maximum runtime. In the latter case, which rarely occurred, the results were checked to ensure that the issue encountered was insensitivity of some parameter(s) in a particular catchment.

Duan, Q. Y., Sorooshian, S. & Gupta, V. K. Optimal use of the SCE-UA global optimization method for calibrating watershed models. *Journal of Hydrology,* **158**, 265-284, doi:10.1016/0022-1694(94)90057-4 (1994).

Parameter	Value
p (number of complexes)	max(5, n)
m (number of points per complex)	2n+1
q (number of points per sub-complex)	n+1
alpha (nb. of evolution steps for each sub-complex)	3
beta (nb. of successive sub-complexes per complex)	2n+1

Table S7. Shuffled Complex Evolution settings for n calibrated parameters.

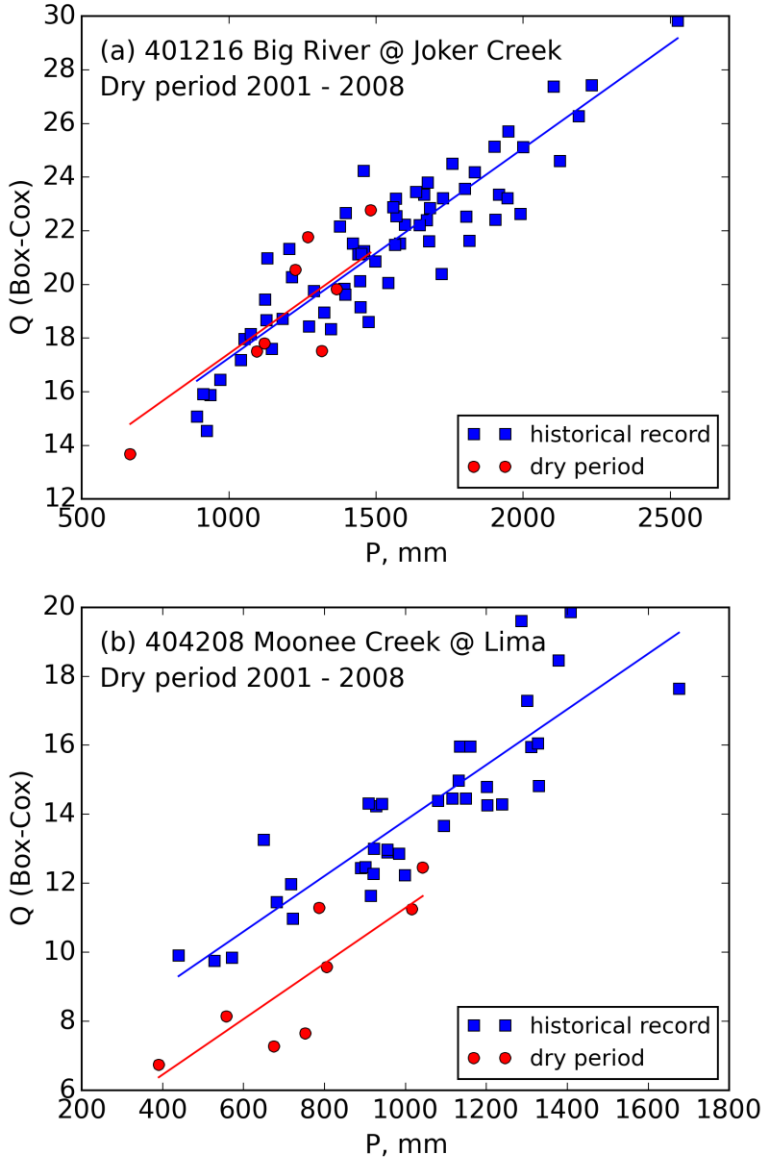

Figure S7. Examples of the stable (a) and shifted (b) rainfall-runoff relationship during the dry period (see Text S2 for the test details).

Text S2. Statistical test for detecting shift in the rainfall-runoff relationship.

Runoff data were modified with the Box-Cox transformation [*Box and Cox, 1964*] to reduce the skewness in runoff distribution and make the rainfall-runoff relationship linear.

The regression model is:

$$Q = a_0 + a_1 \times I + a_2 \times P + e, \qquad (1)$$

where Q is annual runoff, P is annual rainfall, I is a drought indicator, a_0 is the intercept and e is the regression residual. I equals 0 for all years except the Millennium drought (corresponds to the blue color in the Figure S7), and equals 1 for the Millennium drought (corresponds to red color in the Figure S7). A t-test is performed on the drought indicator coefficient a_1, and it determines whether a1 is significantly different from 0. In other words, the test shows whether two regression lines in Figure S7 are significantly different. Statistical testing allowed for both autocorrelation in the data and adjusted the impacts of multiple testing (i.e. testing the same hypothesis on many catchments). For more details see Saft et al. [2015].

Box, G. E., and D. R. Cox (1964), An analysis of transformations, *J. R. Stat. Soc. Ser. B, 26*(2), 211–252.

Saft, M., A. W. Western, L. Zhang, M. C. Peel, and N. J. Potter (2015), The influence of multiyear drought on the annual rainfall-runoff relationship: An Australian perspective, *Water Resour. Res., 51*(4), 2444-2463.

Catchment ID	Observations start	Observations end	% complete	Dry period start	Dry period end
201001	22/05/1947	20/01/2009	99.2	1992	2007
203005	01/06/1943	10/05/2009	85.1	1992	2007
203030	01/06/1969	10/05/2009	99.9	2000	2006
204034	26/10/1951	10/05/2009	92.8	2002	2008
204036	01/01/1952	10/05/2009	99.7	1999	2008
204037	01/07/1952	10/05/2009	83.9	2002	2008
204055	29/02/1972	10/05/2009	98.4	1998	2005
204067	22/04/1983	10/09/2008	100	2000	2006
206014	15/04/1948	25/03/2009	98.3	2001	2007
211013	13/11/1976	10/05/2009	98.8	1993	2006
211014	11/11/1976	10/05/2009	99.1	1993	2006
212040	09/06/1979	29/04/2009	96.9	2002	2008
215002	01/09/1914	10/05/2009	78.8	2001	2008
216004	10/04/1970	24/02/2009	95.0	1993	2008
216009	20/04/1985	10/05/2009	99.6	2002	2008
218001	25/06/1948	10/05/2009	83.7	1994	2008
218007	13/06/1974	10/05/2009	100	1994	2008
219013	12/11/1961	10/05/2009	69.1	1994	2008
219016	18/06/1965	10/05/2009	78.4	1995	2008
219017	08/07/1966	10/05/2009	100	1995	2008
220003	01/09/1966	10/05/2009	100	1995	2008
220004	06/04/1970	10/05/2009	99.3	1994	2008
221002	01/12/1971	16/03/2009	99.1	1995	2006
221010	11/07/1981	10/05/2009	99.2	1995	2006
221201	23/04/1922	31/12/2008	73.2	1995	2008
222004	12/03/1941	22/04/2009	100	1995	2008
222007	25/03/1949	10/05/2009	100	1994	2008
222202	25/04/1922	31/12/2008	80.6	1996	2008
222213	02/08/1957	31/12/2008	96.3	2001	2007
223202	02/06/1947	31/12/2008	98.0	1997	2008
225213	29/06/1963	31/12/2008	95.0	1994	2008
225218	14/04/1967	31/12/2008	98.0	1994	2008
225219	08/04/1967	31/12/2008	96.7	1995	2008
226204	24/10/1924	31/12/2008	90.3	1997	2008
226218	24/06/1955	31/12/2008	85.1	1997	2008
226410	10/07/1953	31/12/2008	83.0	1997	2008
227202	24/06/1955	31/12/2008	97.0	1998	2008
227211	11/01/1957	31/12/2008	98.8	1997	2008
227219	02/04/1966	31/12/2008	98.4	1998	2008
228206	23/05/1959	10/05/2006	99.1	1998	2004
228212	17/03/1962	08/06/2006	99.5	1998	2005
229215	06/03/1968	04/05/2006	96.6	1998	2004
230205	22/06/1955	31/12/2008	96.5	1996	2008
231213	16/04/1959	31/12/2008	96.4	1998	2008

233215	12/10/1956	31/12/2008	97.9	1996	2008
233223	03/06/1970	31/12/2008	95.3	1997	2008
234200	02/04/1957	31/12/2008	97.5	1995	2008
236203	16/06/1920	31/12/2008	83.4	1995	2008
236205	02/08/1948	31/12/2008	98.1	1995	2008
236212	06/06/1965	31/12/2008	99.4	1995	2008
237200	03/06/1948	31/12/2008	97.9	1995	2008
237205	12/01/1963	31/12/2008	98.2	1995	2008
237206	29/02/1964	31/12/2008	99.1	1995	2008
401210	12/10/1932	31/12/2008	100	2002	2008
401215	19/10/1929	31/12/2008	100	2002	2008
401216	14/11/1934	31/12/2008	99.9	2001	2008
401217	27/07/1971	31/12/2008	99.9	2002	2008
403217	10/08/1962	31/12/2008	100	2002	2008
404208	26/01/1963	31/12/2008	100	2001	2008
405205	17/06/1939	31/12/2008	100	1995	2008
405209	13/12/1945	31/12/2008	99.9	1995	2008
405212	22/11/1945	31/12/2008	83.8	2002	2008
405214	07/02/1947	31/12/2008	100	2002	2008
405215	08/02/1947	31/12/2008	70.6	1995	2008
405217	27/03/1954	31/12/2008	100	2001	2008
405219	27/08/1954	31/12/2008	100	1995	2008
405226	11/12/1957	31/12/2008	100	1998	2008
405227	08/02/1958	31/12/2008	99.2	1997	2008
405228	17/09/1958	31/12/2008	99.9	2001	2008
405229	21/05/1960	31/12/2008	99.8	1996	2008
405230	24/05/1960	31/12/2008	96.1	2001	2008
405231	27/05/1961	31/12/2008	99.8	2002	2008
405240	09/12/1972	31/12/2008	96.9	2002	2008
405241	02/05/1922	31/12/2008	75.9	1994	2008
405245	13/05/1970	31/12/2008	99.9	2001	2008
405246	07/05/1970	31/12/2008	93.7	2001	2008
405248	20/04/1971	31/12/2008	99.9	2001	2008
405251	26/05/1971	31/12/2008	99.9	2001	2008
405274	10/06/1977	31/12/2008	100	2001	2008
405293	12/06/1989	31/12/2008	94.3	2002	2008
406213	02/11/1953	31/12/2008	93.2	2001	2008
406214	11/03/1965	31/12/2008	100	1995	2008
406226	28/06/1978	31/12/2008	99.7	2002	2008
407211	24/09/1943	31/12/2008	45.7	2001	2008
407213	14/08/1943	31/12/2008	96.3	2002	2008
407222	25/08/1955	17/12/2008	99.7	1996	2007
407230	14/05/1963	31/12/2008	100	2002	2008
407239	23/05/1970	31/12/2008	75.3	2002	2008
408206	27/05/1967	31/12/2008	67.4	2001	2008
410024	19/09/1914	10/05/2009	99.9	2001	2008
410026	25/08/1915	10/05/2009	93.2	2001	2008

410038	28/04/1932	10/05/2009	96.9	2001	2008
410044	02/06/1938	16/04/2009	91.3	2002	2008
410047	07/06/1938	03/05/2009	89.5	2001	2008
410048	11/06/1938	23/03/2009	79.6	2002	2008
410057	18/12/1944	10/05/2009	99.0	1994	2008
410061	11/09/1947	10/05/2009	100	2001	2008
410062	01/12/1947	04/03/2009	90.6	1994	2008
410076	07/05/1955	03/03/2009	95.1	1994	2008
410081	14/12/1956	31/12/2005	78.7	1995	2005
410107	20/05/1972	06/05/2009	91.3	2001	2008
410141	21/05/1982	03/03/2009	100	2001	2008
410535	14/06/1958	30/04/2007	91.8	1998	2006
410574	30/06/1972	30/04/2007	97.1	1998	2006
410713	30/03/1957	24/03/2009	100	2001	2008
410731	13/11/1964	10/05/2009	100	2001	2008
411003	28/09/1971	10/05/2009	98.1	2001	2008
415201	02/01/1950	13/12/2008	100	1995	2007
415220	02/02/1963	31/12/2008	97.4	1998	2008
415226	29/04/1971	31/12/2008	100	1995	2008
416003	02/09/1921	24/03/2009	94.8	2000	2008
416008	16/09/1934	10/05/2009	83.8	2002	2008
416016	01/06/1962	22/03/2009	99.6	2002	2008
416020	03/03/1967	23/03/2009	99.1	2002	2008
416023	14/03/1967	24/02/2009	99.4	2002	2008
416039	10/05/1974	10/05/2009	98.3	2002	2008
421018	01/08/1939	10/05/2009	98.6	2002	2008
421026	10/09/1947	10/05/2009	98.0	2001	2007
422319	29/03/1969	21/10/2008	100	1999	2007
422334	18/04/1969	17/10/2008	100	1999	2007
422338	27/10/1972	27/08/2008	100	2000	2007
422350	18/10/1980	26/08/2008	100	2000	2007
422352	21/05/1987	27/08/2008	100	2000	2007
426503	16/03/1969	20/05/2008	99.3	1995	2004

Table S8. Catchment analysis periods and runoff record details. Catchment ID corresponds to that used by the relevant government agency. A record with a high percentage of gaps indicates that the gauge was decommissioned for one, or rarely more, continuous periods of time. PET and rainfall records are complete (no gaps).