Predicting shifts in rainfall-runoff partitioning during multiyear drought: roles of dry period and catchment characteristics.

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

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

[2]CSIRO Land and Water, Canberra, ACT, Australia

* Corresponding author: Margarita Saft, Department of Infrastructure Engineering, The University of Melbourne, Victoria, 3010, Australia

Tel: +61-(0) 3-8344-7305

Email: margarita.saft@unimelb.edu.au

Key points:

- We explain the variability in the magnitudes of shifts in the rainfall-runoff partitioning observed during the decadal Millennium drought
- During decade-long dry periods, the severity of hydrological drought is strongly influenced by the catchment biophysical structure
- Catchments susceptibility to shifts in hydrologic response was mostly related to pre-drought climate and catchment storage characteristics

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version record. Please cite this article as doi:10.1002/2016WR019525.

Abstract

While the majority of hydrological prediction methods assume that observed interannual variability explores the full range of catchment response dynamics, recent cases of prolonged climate drying suggest otherwise. During the ~decade-long Millennium drought in south-eastern Australia significant shifts in hydrologic behaviour were reported. Catchment rainfall-runoff partitioning changed from what was previously encountered during shorter droughts, with significantly less runoff than expected occurring in many catchments. In this article we investigate the variability in the magnitude of shift in rainfall-runoff partitioning observed during the Millennium drought. We re-evaluate a large range of factors suggested to be responsible for the additional runoff reductions. Our results suggest that the shifts were mostly influenced by catchment characteristics related to pre-drought climate (aridity index and rainfall seasonality) and soil and groundwater storage dynamics (pre-drought interannual variability of groundwater storage and mean solum thickness). The shifts were amplified by seasonal rainfall changes during the drought (spring rainfall deficits). We discuss the physical mechanisms that are likely to be associated with these factors. Our results confirm that shifts in the annual rainfall-runoff relationship represent changes in internal catchment functioning, and emphasise the importance of cumulative multiyear changes in the catchment storage for runoff generation. Prolonged drying in some regions can be expected in the future, and our results provide an indication of which catchments characteristics are associated with catchments more susceptible to a shift in their runoff response behaviour.

1. Introduction

Extrapolating from the observed range of catchment responses to new conditions, such as a substantially different climate, is known to be challenging. Yet it is one of the rising issues of contemporary hydrological research and practice, as indicated by the current International Association of Hydrological Sciences (IAHS) decade Panta Rhei [*Montanari et al.*, 2013], which is devoted to improving predictive ability in a changing environment. It is often assumed that it is possible to extrapolate interannual catchment response to longer timescales or that the hydrologic response to longer-term or permanent change can be inferred from past interannual variability. As an example, if a single dry year with a 20% rainfall reduction typically results in a 50% runoff reduction, then a 20% average annual rainfall reduction by 2075 should similarly result in a 50% average runoff reduction, but will it? This assumes a stability or stationarity in catchment functioning that has recently been shown to be invalid [*Chiew et al.*, 2014; *Hughes et al.*, 2012; *Saft et al.*, 2015].

It has been shown that catchment dynamics under long periods of unusual conditions cannot be confidently inferred from shorter periods of similar conditions (i.e. interannual variability), at least in some cases. Australia provides two such examples: the recently experienced decadal drought in the south-east (usually referred as the Millennium drought, circa 1997 – 2008), and the multidecadal drying trend in the south-west. In south-west Western Australia, the cumulative effects of groundwater decline significantly reduced the runoff generation rates, with runoff for a given rainfall being currently much lower than two decades ago [*Hughes et al.*, 2012; *Kinal and Stoneman*, 2012; *Petrone et al.*, 2010]. In south-eastern Australia, a number of studies have reported shifts in the historical rainfall-runoff relationship during the recent decade-long drought, which means that the runoff during this decade was significantly lower than in other similarly dry years of the record [*Chiew et al.*, 2014; *Potter et al.*, 2011; *Saft et al.*, 2015]. These results correspond to a relatively small yet noticeable departure from the Budyko curve during the same period [*Roderick and Farquhar*, 2011], which, as the authors hypothesised, might indicate a change in the coefficient which represents the integral effect of catchment properties.

The recent Millennium drought in south-eastern Australia can be seen as a large-scale natural experiment that provides information with which to explore the interdecadal changes in catchment functioning. This drought has been experienced over a large area covering the Murray-Darling basin and the adjacent east coast of Australia. The vast spatial extent of the drought allows investigation of both catchment influences and drought influences over the reported changes in rainfall-runoff relationships. In earlier studies, the suggested underlying causes of the large runoff declines have been mostly related to the meteorological properties of the drought itself, including higher temperatures [*Cai and Cowan*, 2008; *Potter et al.*, 2011], dominance of autumn rainfall deficits [*Chiew et al.*, 2011; *Verdon-Kidd and Kiem*, 2009], and absence of high rainfall years [*Murphy and Timbal*, 2008]. However, the shifts in the rainfall-runoff average partitioning for a given rainfall have also been associated with catchment characteristics, such as historical climate aridity, slope, percentage of woody cover [*Petheram et al.*, 2011; *Potter et al.*, 2011; *Saft et al.*, 2015]. Comparing catchments which exhibited a shift in behavior during the Millennium drought with those that did not, showed that these two groups generally have statistically significant differences in catchment characteristics, but not in dry period characteristics [*Saft et al.*, 2015], implying that endogenic rather than exogenic causes may be more important for the shifts in the rainfall-runoff partitioning. However, no systematic analysis has been done to date to examine which factors were more important for the shifts in catchment functioning.

Several internal mechanisms have been hypothesised as responsible for the shifts in catchment behaviour, with the strongest suspects being groundwater decline leading to shrinkage of the partial contribution area, and soil-vegetation adaptation/adjustment [*Chiew et al.*, 2014; *Petheram et al.*, 2011; *Saft et al.*, 2015]. In a qualitative sense, vegetation is known to adjust quickly to severe disturbances (such as droughts), so interdecadal changes to species composition, root depth and vegetation cover density are quite likely (e.g *Mueller et al.* [2005]). Groundwater level changes affecting stream-aquifer recharge-discharge rates and directions are also plausible mechanisms potentially responsible for longer-term changes in watershed behaviour [*Brunner et al.*, 2009; *Brunner et al.*, 2011; *Eltahir and Yeh*, 1999; *Peters et al.*, 2003]. Lastly, while soils are relatively conservative in comparison to groundwater and vegetation, some important properties of the soils, such as hydraulic conductivity, water repellence and preferential flow pathways might change on interdecadal scales, especially when coupled with vegetation cover changes (as active depths are influenced by vegetation) [*Angers and Caron*, 1998; *Farmer et al.*, 2003; *Gao et al.*, 2014], and with groundwater storage changes (through partial area contribution changes [*Dunne and Black*, 1970]).

The practical importance of changes in the rainfall-runoff relationship is the impact on streamflow prediction. Current common streamflow prediction methods, i.e. conceptual rainfall-runoff models, are known to produce inaccurate estimates when parameters are transferred between climatically different periods [*Coron et al.*, 2012; *Merz et al.*, 2011; *Vaze et al.*, 2010a]. Challenges in modelling under different climates, and in particular the encountered systematic bias, have been argued to be at least partially caused by changes in catchment functioning [*Peel and Blöschl*, 2011; *Saft et al.*, 2016]. It is not unexpected that the current conceptual models, which rely heavily on calibration, produce highly inaccurate predictions when catchment hydrological response changes. However, since these models are widely used to predict under climatically different conditions, further investigation of the factors related to the changes in catchment functioning is needed, as it can aid in understanding which catchments, or which dry periods, are likely to be modelled with reasonable confidence.

Observed long-term changes in regional climate provide information to explore the issue of instability in catchment functioning further. This article aims to explore what factors are primarily responsible for the observed shifts in rainfall-runoff relationships. A variety of potential factors including historical catchment climate, catchment biophysical properties, and climatic characteristics of the shift period are considered. We view this paper as an exploratory exercise rather than an attempt to find one “true” model. We use information theory and a multimodel inference approach to find (1) how much variance in the rainfall-runoff relationship shifts can be explained, and (2) which factors are more important for explaining the shift in the rainfall-runoff relationship. The systematic analysis of the relative importance of the explanatory factors may provide evidentiary support for the previously suggested hypotheses or lead to the formulation of new hypotheses.

2. Methods

In order to further investigate the sensitivity of rainfall-runoff partitioning to interdecadal changes in rainfall we select a large number of potential influencing factors, and relate them to the size of rainfall-runoff relationship shifts (as defined in section 2.1). Using 116 catchments, we first explore all possible linear regression models with up to 6 (from the initial set of 37) explanatory variables. We repeat similar analyses on subsamples of our catchment set 20 times to check sensitivity to catchment selection. Then we exhaustively explore a subset of 21 better performing explanatory variables to check whether more than 6 explanatory variables can be justified. After each stage the models are compared using several information criteria and conclusions are drawn from the whole set of models (multimodel inference). Please refer to the Figure 3 and sections 2.2 and 2.3 for more details.

2.1. Magnitude of shift

All models are optimised to predict the magnitude of the shift in the annual rainfall-runoff relationship normalised by the expected drought runoff in each catchment. Definition and calculation of the magnitude of shift in the rainfall-runoff relationship follows *Saft et al.* [2015]. Figure 1 illustrates the calculation and meaning of the magnitude of shift in the annual rainfall-runoff relationship. Figure 2 illustrates the distribution of magnitudes of shift in the 116 study catchments located in south-eastern Australia during the Millennium drought. The magnitude of the shift represents the distance between the lower (=dry) part of the historical rainfall-runoff relationship and the rainfall-runoff relationship during the Millennium drought. If the distance between these lines is large, then there was an observable change in the rainfall-runoff partitioning during the Millennium drought compared with the other (typically short) dry periods. The duration of the Millennium drought in the study catchments ranged between 7 and 16 years, being 10.3 years on average. In some catchments (~13%) the Millennium drought was not the only long (⩾7 years) drought found in the record. However, the most common other drought (the WWII drought, present in 9% of the catchments) was less severe in terms of runoff due to high rainfall variability during that drought [*Potter et al.*, 2010]. Due to this and the fact that most of the stations have only one ~decade-long drought, we analyse the Millennium drought versus all other dry years present in the record. The Box-Cox transform is applied to the annual runoff data to linearize the relationship. The magnitude of the shift is calculated for the dry period reference rainfall, which is defined as half of the sum of the minimum and average annual rainfall of the record. The magnitude of shift (M) is calculated as:

$$M(\%) = \frac{Q_{dr}(P') - Q_{exp}(P')}{Q_{exp}(P')} \times 100$$

where Q is transformed back to the normal space, Q_{dr} is the dry period runoff for the dry period reference rainfall (P'), and Q_{exp} is the expected runoff for the dry period reference rainfall. The magnitude of the shift is calculated for each catchment individually. It is important to note here that the magnitude of the shift does not describe the total decrease in runoff during the drought. Rather it only represents the additional reduction in runoff during the Millennium drought compared to similarly dry years in the historical record, expressed as a percentage of the expected drought runoff.

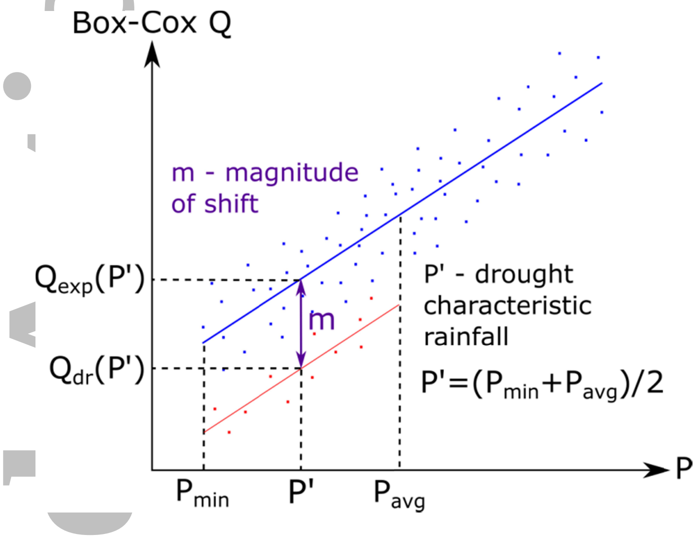


Figure 1. The annual rainfall-runoff relationship and the magnitude of shift (each point represents a year in the time series: red – years of the long drought (≥7 years), blue – other years).

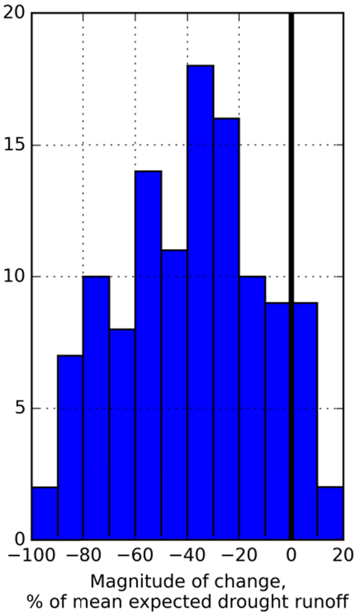


Figure 2. Histogram of the magnitudes of shift in the rainfall-runoff relationship during the Millennium drought in the study catchments.

2.2. Multimodel inference

The multimodel inference approach was chosen to aid investigation of complex and possibly interrelated physical processes with the limited data available. Many of the mechanisms potentially responsible for shifts in the rainfall-runoff relationship are hard to measure directly. Nevertheless, proxies and indicators are available. A challenge is that there is often more than one potential proxy (explanatory variable) that can be linked to a given mechanism. Consequently, variables are typically correlated. Also, some of the variables (e.g. aridity index) can be related to multiple potential mechanisms, which does not allow direct interpretation. Lastly, interactions between variables might aid insight. Based on these considerations, there are likely to be multiple similarly good models, and the "best" model cannot be expected to be selected by formal criteria with confidence. The multimodel inference approach somewhat reduces the impact of these issues.

We employ linear regression for three main reasons. First, we prioritise having an extensive set of potential predictors (proxies) over variety in regression form (for example, linear, log-linear, power). Second, we do not have a priori information on more suitable forms for the regression model. And third, there are positive and negative values in both dependent and independent variables, which excludes using some of the possible regression forms (e.g. log-linear, exponential) without further data transformation. A potential drawback of linear regression is its sensitivity to outliers, especially at the ends of the variable ranges. To address this issue we perform subsample testing. We also use statistical tests (D'Agostino [*d'Agostino*, 1971; *d'Agostino and Pearson*, 1973] and Anderson-Darling [*Stephens*, 1974]) to assess the normality of the residuals. Generally, correlation between independent variables (collinearity) is problematic for linear regression, but in the multimodel inference approach the information criteria used penalise additional regression terms that contribute little additional information. Assessing the statistical significance of the regression coefficients in the resulting best models also addresses this issue.

The main advantage of the brute force approach over stepwise regression based methods is the default preservation of any interplay between predictors. Pure stepwise regression is likely to miss the information from interacting predictors if two (or more) predictors only provide information together, but appear to be non-informative when considered individually. The robustness of the brute force approach to the uncertainty in the appropriate predictors comes at a cost of high computing resource demand. On the other hand, models having too many predictors in comparison to the available data will inevitably overfit the available data.

Given that we have 37 predictor variables, we approached the model inference task in several stages to reduce the computational burden (see Figure 3). The initial analysis (main search) provided the main set of results, and two subsequent checks were conducted to assess whether these results were robust. The main search considered all possible models with up to 6 predictors from a total set of 37 predictors. This allowed us to detect interactions between the predictors while avoiding models that are likely to be overfitted. In order to investigate the stability of these results we run subsample testing (Check 1). Here the catchment set is randomly divided into fifths, and each fifth is left out in turn while the remaining catchments are used to fit the models similarly to the main search (37 predictors, and maximum 6 predictors per model are allowed). Subsample testing was repeated four times, thus the models were fitted to 80% of catchments 20 times, which resulted in 56,703,980 optimised regression models. The preferred models and predictors are compared internally and to the results of the main search. Based on the results from the main search and the subsample testing we select better performing predictors for further consideration and run an exhaustive search (Check 2) of all possible models with up to 21 predictors per model. These 21 predictors include all predictors present in the best models for the main search and subsample testing, and predictors which were present in at least 10% out of the 50 top models according to each information criteria used (see the section below) for the main search, and predictors which were present in at least 25% out of the 50 top models (*20 repetitions *3 information criteria) for subsample testing. The exhaustive search enabled us to check whether more predictors per model should be allowed, i.e. whether the fit can be significantly improved with more than 6 predictors.

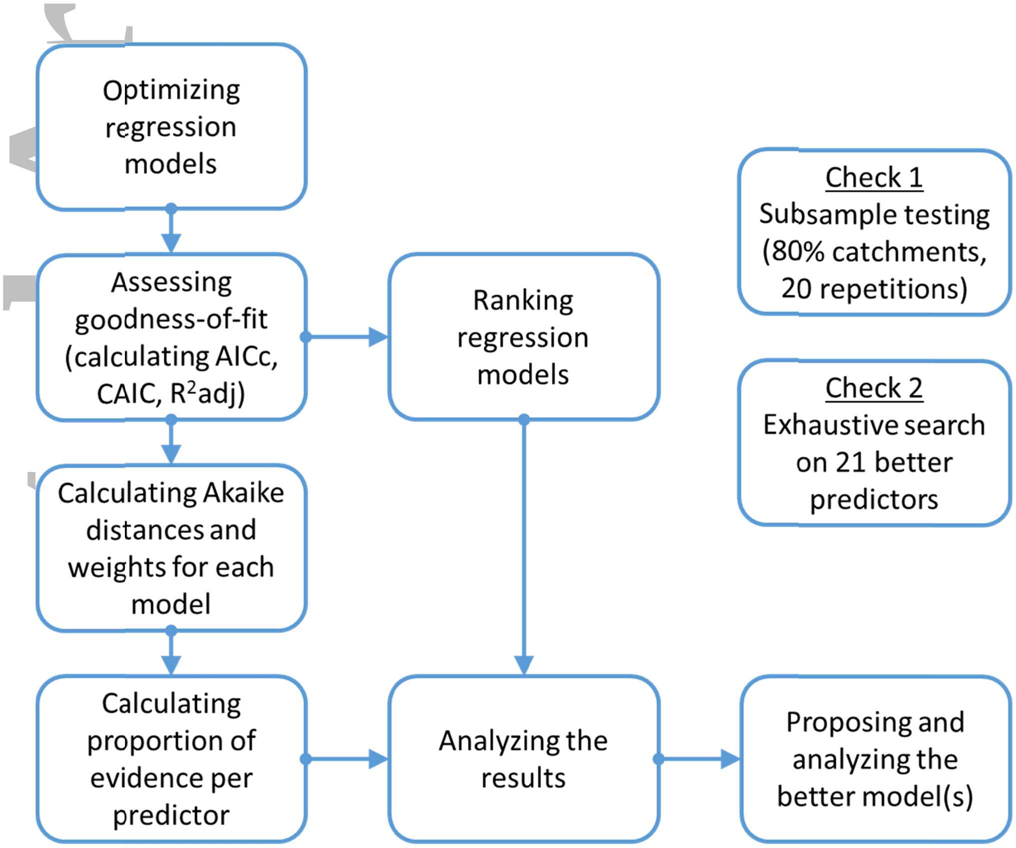


Figure 3. Analyzis steps.

2.3. Information criteria

We assess and compare the regression models with information criteria. There are a number of criteria that can be used for this purpose, but all are based on a term that represents goodness of model fit and a penalty term for extra variables to avoid over-fitting. Depending on the penalty term used, the criteria will either favour simpler models, even if they do not fit well, or more comprehensive models, at the cost of likely over-fitting. We employ three criteria to compare the models: adjusted R^2 (R^2_{adj}), Consistent Akaike Information Criterion (AICc), and Complete Akaike information criterion (CAIC). R^2_{adj} is based on the residual sum of squares, whereas AICc and CAIC are based on maximised log-likelihood. These criteria also range in their tendency to choose potentially under- or over-fitted models. In our case, adjusted R^2 penalises over-fitting the least, AICc moderately, and CAIC heavily. This diversity gives us an expectation that different criteria will select different models as best. However, with real world data we have no indication of the true fit, and we are not sure whether any particular model is under-fitted or over-fitted. Therefore we employ several criteria, as it gives information on how the selected criteria influence model ranking. The particular formulas are:

$\mathrm{AIC} = -2 * llf + 2 * k$ [*Akaike*, 1973]
$\mathrm{AICc} = \mathrm{AIC} + (2 * (k + 1) * (k + 2)) / (n - k - 2)$ [*Hurvich and Tsai*, 1989]
$\mathrm{CAIC} = -2 * llf + k * (\ln(n) + 1)$ [*Bozdogan*, 1987]
$R^2_{adj} = 1 - [(n - 1) / (n - k - 1)] [1 - R^2]$ [*Ezekiel*, 1929; *Wang and Thompson*, 2007]

where *llf* is the log-likelihood function, *k* is the dimension of the model, and *n* is the number of observations (in this case catchments). The likelihood function is a measure of probability of the model given the data, and the log-likelihood is used to find the maximum likelihood efficiently (for further details see [*Beven* [2010]; *Montgomery and Peck* [1992]]).

Akaike based criteria are uninterpretable on their own, but they allow ranking of models from best to worst, and also calculation of Akaike distances, Akaike weights, and finally the proportion of evidence for individual predictors. The latter is the main focus of this study and follows the methodology as presented in *Burnham and Anderson* [2002].

To rank and compare the models, we employ AIC differences (though we calculate them on AICc and CAIC instead of the original AIC).

$\Delta_i = \mathrm{AIC}_i - \mathrm{AIC}_{min}$, where AIC_{min} is the AIC of the best model.

Further, we calculate model probabilities, or model weights (which for the whole set will add to 1):

$W_i = \exp(-0.5* \Delta_i) / \sum \exp(-0.5* \Delta_{all})$.

Comparison of model weights allows determination of how much better one model is than another (i.e. W_1 / W_2 is termed the evidence ratio).

Finally based on the model weights, we calculate the proportion of evidence for each predictor, as the sum of the weights of the models which include this predictor.

2.4. Independent variables (predictors)

As processes potentially responsible for long-term change are not directly measured (or at least not for a sufficient amount of time / in sufficient detail), we have to rely on measurable proxies. We employed 37 potential predictors, comprising three groups: catchment historic (pre-drought) hydroclimatic characteristics, catchment physical properties, and climatic characteristics of the dry period. Table 1 summarises the variables included in the analysis. Historic hydroclimatic characteristics and dry period climate characteristics were calculated from the rainfall, runoff, and potential evapotranspiration (PET) records. Aridity index was calculated as rainfall / PET, following *UNEP* [1992]. Note that particular attention is paid to the characterisation of historical low-flow dynamics. In this case some low flow characteristics are used as a proxy (indicator) of groundwater variability. In particular, following ideas of *Brutsaert* [2008], who demonstrated that groundwater storage change averaged for a catchment can be inferred from base flows, we calculate several statistics based on annual 7-day minimal flows, and use those to get an indication of historical groundwater behaviour. We have to resort to using low flow variability in place of groundwater storage variability, as suitable groundwater data were unavailable for our catchments.

Some of the variables used are proxies for the same property or process, and therefore they are closely related. Figure S1 in the Supplementary Information illustrates the correlation matrix, and highlights the interdependencies. As noted before, the correlation between predictors will likely result in multiple models having similar performance, and possibly in a similar proportion of evidence for closely related predictors. We will pay attention to the cross correlation issue in the analysis of the results later.

Table 1. Independent variables used for the main search and subsample testing (variables also used for the exhaustive search are **in bold).**

Catchment pre-drought hydroclimatic characteristics (20)	Catchment physical properties (9)	Dry period climate characteristics (8)
• **Average annual rainfall** • **Aridity index (rainfall/PET)** • **Average annual runoff** • **Average maximum daily temperature** • **Average minimum daily temperature** • **Coefficient of variation (C_v) of annual runoff** • **C_v of annual rainfall** • **C_v of monthly rainfall** • **C_v of annual runoff ratio** • **C_v of annual minimal 7-day flows** • Average annual PET • Runoff ratio • Baseflow index (BFI) • % cease to flow • Average annual minimal 7-day flow • Range of annual minimal 7-day flows • C_v of annual PET • C_v of annual BFI • C_v of monthly runoff • Annual minimal 7-day flow divided by mean 7-day flow	• **Catchment area** • **Mean elevation** • **Elevation range** • **Stream density** • **Mean solum thickness** • **Percentage of woody cover** • Mean plant available water capacity • Mean slope • Stream length	• **Drought rainfall anomaly** • **Drought length (>7 years)** • **Drought anomaly of winter rainfall** • **Drought anomaly of spring rainfall** • **Drought anomaly of summer rainfall** • Drought anomaly of autumn rainfall • Drought average maximum daily temperature anomaly • Drought average minimum daily temperature anomaly

2.5. Dataset

Runoff data from 116 catchments in south-eastern Australia were used for this study. The catchment set was inherited from *Saft et al.* [2015], which in turn used catchments selected for the 'Murray-Darling Basin Sustainable Yields' [*Chiew et al.*, 2008] and 'South Eastern Australian Climate Initiative' projects [*CSIRO*, 2012]. The original set consisted of 228 catchments and was developed to include unimpaired catchments without flow regulation or alteration, and only minimal, if any, urbanisation and forestry [*Vaze et al.*, 2010b]. Despite care being taken during catchment selection in initial dataset construction, we cannot fully rule out anthropogenic influences. Furthermore, potential impacts from adjacent areas (e.g. groundwater extraction in neighbouring catchments [*Hisdal et al.*, 2001]), which may influence within-catchment processes, were not considered. 124 of these catchments were found to have a severe, extended, and uninterrupted dry period (the Millennium drought) circa late 1990s and 2000s (for details on the dry period extraction algorithm see [*Saft et al.*, 2015]). For the current study, 8 catchments were excluded: 5 were excluded as they did not have a value for one of the variables (coefficient of variation of annual 7 day minimal flow) due to their stable ephemeral regime; and 3 were excluded because of having high positive magnitudes of shift in the annual rainfall-runoff relationship (>+40% of expected dry period flow). Of these three catchments, one is highly ephemeral, another has suspected data issues during the dry period and the other includes a very large flood at the end of the dry period despite satisfying the annual rainfall moving window dry period criteria. We acknowledge that there might be some meaningful information in these points, but removed them because they may have an undue influence on the regression given they are outlying data points near the extremities of the variable range. A list of the stations used along with details on the start and end of the observation record, start and end of the dry periods, and % data completeness can be found as table S8 in *Saft et al.* [2016] (8 gauges which were excluded from this study are 225213, 233223, 405293, 410081, 415220, 415226, 422338, 426503). The average record length in the study catchments is 52.5 years, and the average data completeness is 94.7%. The map of the catchments with the magnitudes of shift in the rainfall-runoff relationships is shown in Figure 4.

Meteorological data used in this study (daily rainfall, PET and temperature) were extracted from the SILO Data Drill (www.longpaddock.qld.gov.au/silo, *Jeffrey et al.* [2001]). Catchment physical properties were obtained as part of the parent dataset [*Vaze et al.*, 2010b], and were originally estimated from a DEM (grid resolution of 20 and 25 m), interpretation of aerial photography (Bureau of Rural Sciences, Australia), and Australia Soil Resource Information System.

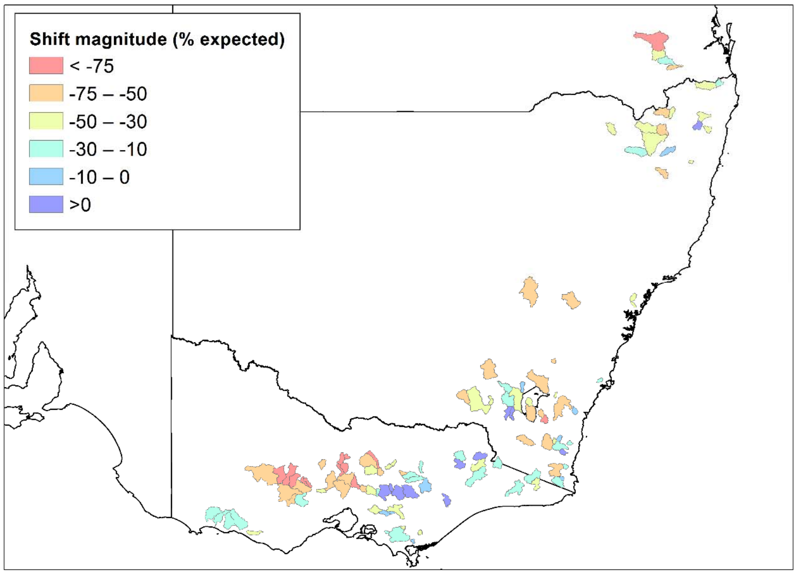


Figure 4. Map of the study catchments with corresponding magnitudes of shift in rainfall-runoff relationship.

3. Results

3.1. How much variance in rainfall-runoff relationship shifts can we explain?

The proportion of variance explained by the tested models is usually assessed through the coefficient of determination (R^2). As we employ multiple linear regression with a varying number of predictors, we use the adjusted R^2 (R^2_{adj}), which penalises for extra predictors and is therefore more appropriate for comparing models of different complexity. The highest R^2_{adj} for the main search was 0.65 which translates to 65% of variance explained by a 6 predictor model. For the subsample testing (Check 1) R^2_{adj} ranged between 0.612 - 0.701, while mean and median R^2_{adj} were both 0.657. For the exhaustive search (Check 2) R^2_{adj} was 0.665 which corresponds to 66.5% of variance explained by the model based on 11 predictors. From this, the increase in the number of predictors from 6 to 11 only improves the model fit marginally (by 1.5%). Therefore the results of the exhaustive search (Check 2) indicate that the results from the main search are likely to be robust. The issue of the optimal number of parameters (i.e. model complexity) will be discussed further in the section 3.3. Figure 5 shows R^2_{adj} for the 50 best models for both the main and exhaustive searches. Overall, we can explain approximately 65% of the variance in the magnitude of shift of the rainfall-runoff relationship during the Millennium drought.

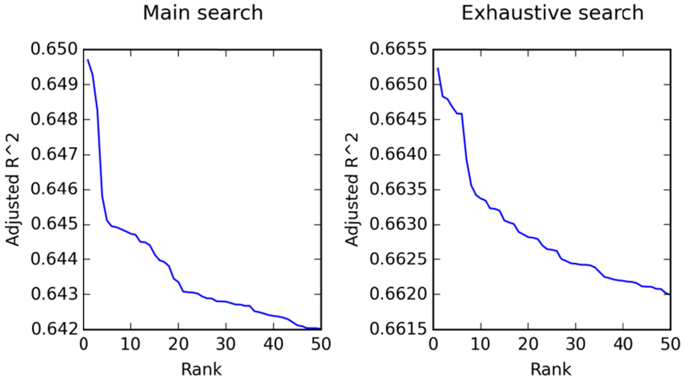


Figure 5. Adjusted R2 for 50 models ranked in descending order.

3.2. Proportion of evidence for individual predictors

According to the proportion of evidence allocated to individual predictors, 7 out of our 37 explanatory variables clearly stood out (Figure 6, Figure 7). The proportion of evidence for a given predictor is based on the sum of the Akaike weights for models that included that predictor, with higher Akaike weights representing better models in the set. Figure 6 shows the proportion of evidence (ranges from 0 to 1) for each predictor in order of descending importance. The proportion of evidence presented here were calculated for both AICc and CAIC. In particular, pre-drought climate aridity index (indicator of historical catchment wetness) and pre-drought coefficient of variation of annual minimal 7-day flow (a proxy for groundwater level variability) had a proportion of evidence of ~0.9, suggesting they are the most influential predictors. Drought anomaly of spring rainfall, and pre-drought coefficient of variation of monthly rainfall had proportions of evidence close to 0.6, and the last three (mean solum thickness, pre-drought coefficient of variation of annual rainfall, and drought rainfall anomaly) had proportions of evidence ~0.3. All the other variables had allocated proportion of evidence close to 0.1 or less. Results of the subsample testing (Figure 7) support the finding from the main search, although they show some variability (here results obtained with AICc are presented, but CAIC results are nearly identical). Importantly, variable preferences obtained from the AICc and CAIC criteria are in very good agreement, even though these criteria have a clear distinction in the preferred model complexity, and therefore differ in models chosen as “best” or “better”. Moreover, the same predictors frequently appear in the best models when the models are ranked according to R^2_{adj} (not presented here).

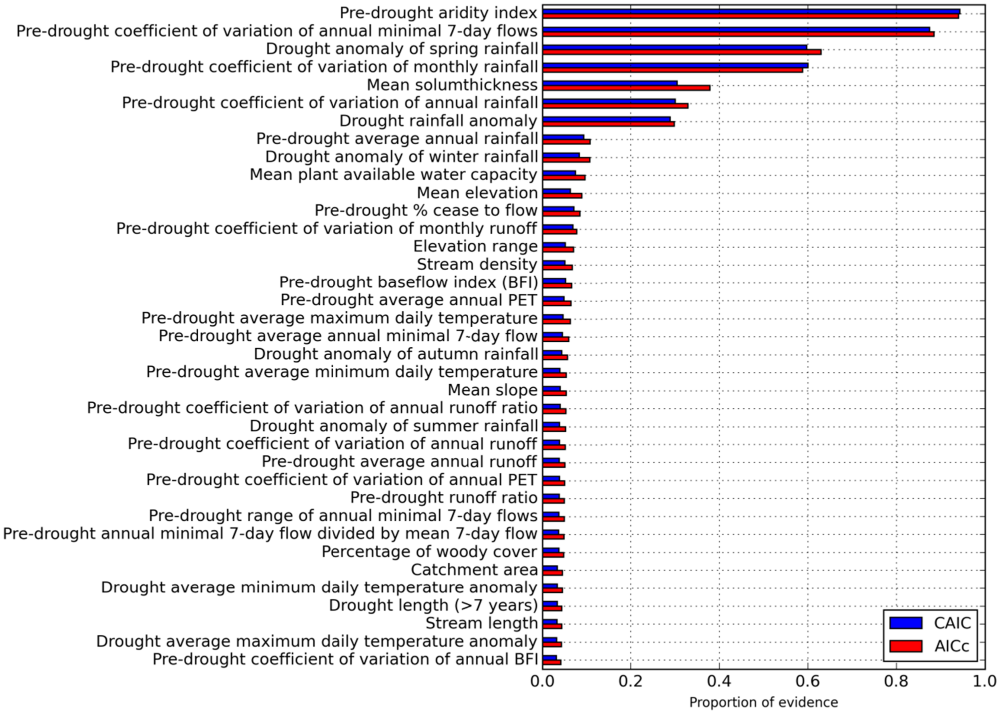


Figure 6. Proportion of evidence for the individual predictors (sorted by proportion of evidence according to AICc in descending order).

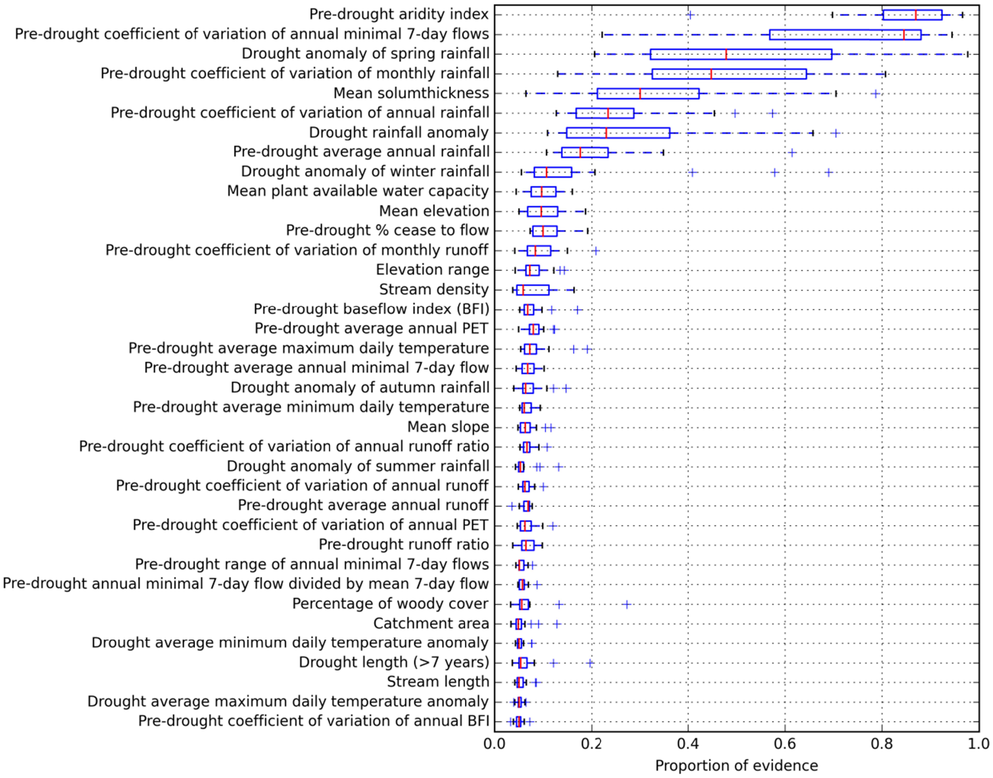


Figure 7. Results of the subsample testing: proportion of evidence for the individual predictors (in the same order as in Figure 6)

3.3. Exploring the “better” models

As discussed earlier, no clearly best model emerged from the model ranking, i.e. the evidence ratios between the best ranked model and following model(s) are small. For example, the best model according to AICc is only 1.07 times better than the 2nd best model, and 1.27 times better than the 5th best model. As the evidentiary support for any particular model is relatively weak, model selection uncertainty is likely to be high, and results are likely to vary between independent samples [*Burnham and Anderson*, 2002].

There are two good starting points to suggest a better model for approximating the rainfall-runoff relationship shift. Firstly, the proportion of evidence for individual predictors (see section 3.2) provides a candidate list of the 7 most influential predictors to be included. Secondly, the exhaustive search results can provide some indication on the optimal number of predictors. Figure 8 shows the results for the best five models for each number of predictors

(5 best single predictor models, 5 best models containing two predictors, etc.). Even though the formal maximum (best fit with number of predictors already factored in) is reached with 7 and 11 variable models according to the AICc and R^2_{adj} criteria respectively, the increase in fitting performance plateaus after 4-5 predictors. According to CAIC, the best models are 4 predictor models, though CAIC sometimes tends to pick underfitted models [*Anderson et al.*, 1998].

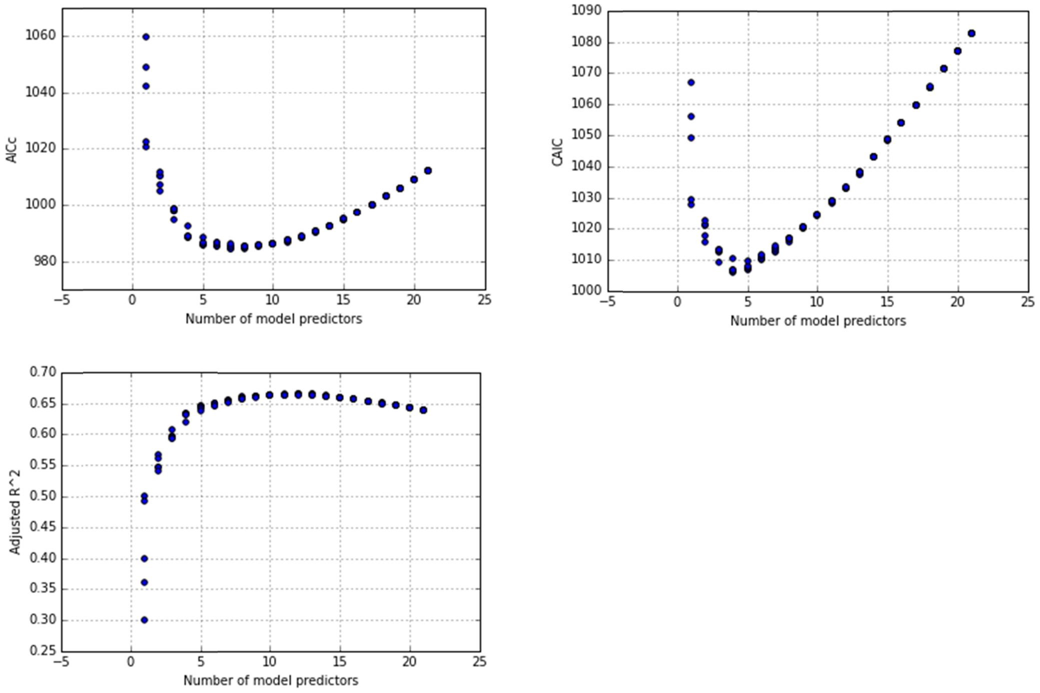


Figure 8. AICc, CAIC, and adjusted R2 for the 5 best models of varying complexity (number of predictors)

Having the initial ranked list of the 7 most influential predictors, and the suspicion that the number of important predictors can be reduced to 4 or 5, we analyse the coefficients of the potential models and their statistical significance. We start with the 4 most influential factors: the pre-drought climate aridity index (rainfall/PET), the pre-drought coefficient of variation of annual minimal 7-day flow, the pre-drought coefficient of variation of monthly rainfall, and the drought anomaly of spring rainfall. The resulting regression model results in R^2_{adj} of 0.63, and all of the regression coefficients are significant (see

). Based on Figure 6 there are 3 variables that potentially may be appropriate to add as a fifth variable to the four variable model. When we added the pre-drought coefficient of variation of annual rainfall or the drought rainfall anomaly as a predictor to the model, the added variable was not statistically significant, and a previously significant coefficient (the pre-drought coefficient of variation of monthly rainfall or the anomaly of spring rainfall respectively) also became insignificant. This suggests that a high proportion of information is shared in these predictor pairs, which is also reflected by the correlation between factors. Pearson's r for pre-drought coefficients of variation of annual and monthly rainfall is 0.81, and for anomalies of spring or overall drought rainfall it is 0.67. However, adding mean solum thickness as an extra variable results in all regression coefficients being significant, including the mean solum thickness itself.

Table 2. Regression coefficients and their statistical significance for selected models

Predictor	4-predictor model		5-predictor model	
	Regression coefficient	Significance (p)	Regression coefficient	Significance (p)
Pre-drought climate aridity index (rainfall/PET)	65.9	0.000	70.0	0.000
Pre-drought coefficient of variation of annual minimal 7-day flow	-7.41	0.001	-6.90	0.002
Pre-drought coefficient of variation of monthly rainfall	38.0	0.001	34.3	0.002
Drought anomaly of spring rainfall	0.44	0.004	0.44	0.004
Mean solum thickness	-	-	-14.6	0.030

Regression coefficients represent the relationship between the magnitude of the shift in the rainfall runoff relationship during the Millennium drought (note, it typically has negative values) and the predictors. The signs of the coefficients indicate that runoff reductions were exacerbated in drier catchments, in catchments with higher interannual variability of the groundwater storage, with less variable monthly rainfall and also where drought spring rainfall anomalies were more prominent. Additionally, catchments with deeper soils also tended to have less runoff than historical elasticity of rainfall on runoff suggested. To explain this interpretation further, the pre-drought aridity index (note that it is defined as rainfall/PET) and coefficient of variation of monthly rainfall are positively correlated with the magnitude of shift. This means that the rainfall-runoff relationship shifts further downward with decrease in these predictors (i.e. higher aridity, less variable rainfall). The pre-drought coefficient of variation of annual minimal 7-day flow and the mean solum thickness are inversely correlated with the magnitude of shift, therefore the rainfall-runoff relationship shifts further downward with an increase in these predictors (more variable storage, deeper soils). Drought anomaly of spring rainfall (usually a negative value) is positively correlated with the magnitude of the shifts; hence the bigger the reduction in spring rainfall the greater the downward shift in the rainfall-runoff relationship.

4. Discussion

There was no single stand out model in our study, but there were predictors which clearly had more explanatory information than others. The fact that essentially the same predictors were identified as important by three information criteria with different preferences for model complexity gives some confidence that these results are meaningful and robust. The seven most informative predictors are (in order of decreasing importance): pre-drought climate aridity index, pre-drought coefficient of variation of annual minimal 7-day flow, drought anomaly of spring rainfall, pre-drought coefficient of variation of monthly rainfall, mean solum thickness, pre-drought coefficient of variation of annual rainfall, and drought rainfall anomaly. These predictors may directly indicate the physical mechanisms of shifts in catchment response, or just indicate the kind of catchments that are more vulnerable to protracted drought stress without pointing towards a particular mechanism. Some of the best variables are correlated (refer to Figure S1 in the Supplementary Information). However, surprisingly, strong correlation between predictors does not necessary mean that the proportion of evidence allocated to those factors is similar. On the contrary, it can be quite different (e.g. pre-drought aridity index and annual rainfall have ~0.95 Pearson's r, but proportions of evidence from Akaike based criteria are >0.9 and ~0.1 respectively).

Historical aridity index is shown to be very closely related to the magnitude of the shift in the rainfall-runoff relationship. Aridity is known to be a major factor defining long-term water balance (see [*Budyko*, 1974] and numerous related studies, e.g. *Gudmundsson et al.* [2016]), as well as catchment sensitivity to interannual changes in rainfall [*Berghuijs et al.*, 2014] / historical elasticity [*Chiew et al.*, 2006]). Our study shows that aridity is also a key factor influencing sensitivity to interdecadal changes, i.e. departure from the historical rainfall-runoff relationship. Therefore catchment aridity influences mean annual water balance, elasticity and changes in elasticity over ~decadal timescales. Many landscape characteristics are also distributed along the aridity gradient, so in some sense aridity is a characteristic which integrates a number of potentially relevant catchment properties. Therefore it is not surprising that the aridity index is related to both catchment rainfall-runoff partitioning and its susceptibility to change. Interestingly, *Van Loon and Laaha* [2015] found that climate wetness (average annual rainfall) was related to the runoff anomaly of shorter (typically 0.5-2 months) droughts and wetter catchments were associated with larger absolute runoff anomalies during drought. In our study, absolute runoff anomalies also tended to be higher in high rainfall catchments, but wetter catchments were associated with smaller shift magnitudes (i.e. they showed less susceptibility to change in the runoff response between long and short droughts). The increase in air temperature during the drought was not related to shifts in catchment rainfall-runoff relationships, in contrast to earlier studies (e.g. *Cai and Cowan* [2008]) that indicated a rise in temperature played a role in the hydrologic impact of the Millennium drought.

Variability of minimal 7-day annual flow is the second most important factor in defining the magnitude of shift in the annual rainfall-runoff relationship. This means that interannual groundwater variability is a key factor of catchment vulnerability to persistent change. The negative regression coefficient indicates that an increase in interannual groundwater variability leads to a greater shift in the rainfall runoff relationship under extended drought. This result is partially counterintuitive: if the groundwater system is very responsive, then a single dry year is likely to induce notable groundwater decline and possibly change the surface-groundwater connection state [*Tallaksen and Van Lanen*, 2004]. This suggests that single and multiyear responses could be similar, but here we observed they were not. The different response to single and multiyear drought suggests a cumulative change in groundwater connection, which is in line with the earlier research [*Eltahir and Yeh*, 1999; *Peters et al.*, 2003]. We suspect that catchments with more interannually variable groundwater either have greater potential to deplete the groundwater storage over several years, or increase the disconnected area, or a combination of both. The idea of prolonged transition in connection state in some catchments is consistent with the observed widespread groundwater decline during the Millennium Drought [*Chen et al.*, 2016; *Leblanc et al.*, 2009].

We found that soil depth was related to the magnitude of shift. All other factors kept equal, deeper soils resulted in a more pronounced change in the rainfall-runoff relationship. A number of studies have argued that the seasonal soil moisture dynamic covers the extent of soil storage, i.e. soil typically fully dries during the dry season, and then fully saturates during the wet season [*Western et al.*, 2002]. Moreover, only the top layer is often perceived as important for storm runoff generation. However, deeper soil might need more time to respond, particularly in catchments with large soil water storage capacity. Modelled longer-term soil moisture dynamics during the Millennium drought based on the Global Land Data Assimilation System (GLDAS-2) have been reported: the top layer of soils (0-10cm) did not show any trend behaviour since 2002, while the lower level (1-2m deep) showed a continuous decline until 2008 [*Loeb et al.*, 2016]. Possibly it is only when climate shifts for an extended time that the impact of the wetness of deeper layers of soil on streamflow generation is exhibited. Another reason might be that soil depth relates to root depth, and it is the latter that is actually important.

One previously suggested mechanism, relating to partial area contribution [*Dunne and Black*, 1970; *Petheram et al.*, 2011], may explain the presence of the aridity index, groundwater variability and soil depth factors in our list of top variables. The partial area contribution mechanism suggests that the area where groundwater is perched close to the surface needs relatively little rainfall to reach saturation (low initial loss) and hence it is ready to convert most of the rainfall to runoff. When the groundwater level falls, this area shrinks, and then a higher initial loss is needed to produce runoff through the saturation excess mechanism. For the Millennium drought the hypothesised importance of the combination of shallow groundwater and seasonally or temporarily organised soil patterns was initially suggested by *Petheram et al.* [2011], who hypothesised that pre-drought groundwater levels close to the surface amplified the surface runoff, and this effect was diminished during the drought. In Western Australia, this mechanism has also been explored based on the simultaneous analysis of streamflow, bore groundwater levels and stream salinity data [*Kinal and Stoneman*, 2012]. It was found that even though groundwater discharge to the stream was minor and occurred during high rainfall events, the impact of falling groundwater on runoff ratios was disproportionally large, which was attributed to the indirect streamflow facilitation mechanisms such as discussed above. The fact that more variable groundwater storage and deeper soils are factors strongly related to the magnitude of change in the rainfall-runoff relationship is consistent with this mechanism.

We found a high importance of spring rainfall decline in explaining shifts in the rainfall-runoff relationship, unlike other studies, which have claimed that large autumn or autumn-winter rainfall deficits are responsible for amplified runoff decline. In our results autumn rainfall deficits did not make it to the exhaustive search, i.e. it was one of the clearly unpromising factors. There is consensus on the idea that large autumn deficits are a characteristic feature of the Millennium drought. However, evidence for this impacting change in rainfall-runoff response is contradictory. In the modelling study of *Potter and Chiew* [2011], changed seasonality accounted for 11% of runoff reduction, and they argued that larger autumn deficits are important. Conversely, *Saft et al.* [2015] did not find differences between autumn anomalies in catchments with different shift magnitudes. In our catchment set there is a variety of seasonal rainfall regimes, including winter-dominated and summer-dominated catchments. However, in many catchments the runoff season is winter and early spring. In summer evaporative demand is high, so catchments stay dry, therefore there is a large initial loss. Over autumn and winter catchments become wetter, soil storage “fills up” (see fill and spill hypothesis [*Tromp-Van Meerveld and McDonnell*, 2006]), soil moisture patterns eventually become organised and within-catchment connection increases [*Western et al.*, 2001]. When this occurs, runoff is generated. The importance of spring rainfall anomaly is likely to be related to it corresponding to the end of the main runoff production season. Thus the role of the autumn deficits is likely to be indirect in that they may delay the time when catchments became sufficiently wet and achieve the state of connection towards spring, which then allows the spring rainfall anomalies to have a bigger impact on runoff generation.

The importance of the historical monthly rainfall variability is largely related to seasonality. Less seasonal catchments were more likely to have a shift in the rainfall-runoff relationship. Catchments with a prominent wet season have greater opportunity to develop organised soil moisture patterns, especially when out of phase with evaporative demand. Less seasonal catchments might be more vulnerable to soil moisture pattern disconnection, and thus shifted behaviour.

This study was limited by adopting simplified parametric representations of complex and possibly interrelated physical processes. Better representation of catchment storage dynamics and vegetation might give more insight, but we are limited by the data availability from semi-natural catchments with long-term records. Geology and land use are expected to be important factors influencing drought propagation through the catchment systems [*Stoelzle et al.*, 2014; *Van Lanen et al.*, 2013; *Van Loon*, 2015; *Van Loon and Laaha*, 2015]. Regarding land use, we addressed this partially by the initial catchment selection process and partially by investigating the influence of the percentage of woody cover, therefore separating grassland (mostly pasture) and forest land uses. We found that the percentage of woody cover was not an important predictor in this study, which indicates these land use factors were of limited importance. Nevertheless, it is nearly impossible to avoid human disturbance completely, and hydrologically important land use and land cover changes might have appeared gradually or at some point during the analysis period. To gain some insight in this issue, we analysed pre-1997 records of the residuals from the rainfall-runoff relationships as a part of the preliminary analysis undertaken (not presented here), and no significant trend was found in most of the catchments, which gives us some confidence that the historical rainfall-runoff relationships were relatively stable in most of the catchments. Human impacts during the Millennium drought (such as increased water demand or government-induced limitations of water usage) are likely to be present at least in some of the catchments, however, it is not known how these impacts compare to the human response to the other historical dry periods, and they were not accounted for in this study. Regarding the geology, this study was limited to using proxies of groundwater storage; it did not include any predictors directly representing catchment geological structure. The previous drought propagation research indicated that low flow metrics such as baseflow index (BFI) can be seen as (and possibly used in place of) a combination of a number of geological metrics, each of which was found to have a relatively low information content compared to the BFI [*Van Loon and Laaha*, 2015]. In our study we employed a range of pre-drought low flow metrics to consider geological controls over long-term changes in storage and subsequently runoff. Lastly, we used a simple linear regression, and it is possible that non-additive and non-linear forms might give better fit to the data, given the complexity of the processes involved.

A theoretical framework for considering the shifts in the rainfall-runoff relationship is the catchment-climate co-evolution thesis where catchment properties, including catchment morphology, soils, groundwater-surface water interactions, dominant runoff pathways, channel networks, and vegetation co-evolve in response to the climate driving [*Troch et al.*, 2015; *Wagener et al.*, 2010]. While most evidence of catchment co-evolution comes from paleogeography, i.e. very long timescales, observed changes in rainfall-runoff relationships over much shorter time, in order of ~decade(s), suggest that hydrologically important catchment properties can change in the historical record. Our results imply that catchment storage properties (i.e. variability of groundwater storage, soil depth) and initial catchment climate (i.e. aridity, seasonality) define catchment readiness to change in response to climate drying.

This paper demonstrates that during decade-long dry periods, the severity of hydrological drought is strongly influenced by the catchment biophysical structure. This influence indicates that changes in catchment processes occurred in some catchments during decade-long drought. Therefore, conventional statistical methods and rainfall-runoff modelling based on observations from non-shift periods may result in overestimation of water availability in a drier climate.

5. Conclusions

This study is focused on finding the factors which can explain the magnitudes of shift in the rainfall-runoff partitioning observed during the Millennium drought in south-eastern Australia. These magnitudes of shift describe the difference between runoff response to reduced rainfall for the recent ~decade-long Millennium drought and the other dry years of the record. The presence of such shifts in catchment behaviour indicates the potential for instability in catchment functioning due to climatic variability and change. This instability presents a challenge to typical hydrological methods which implicitly assume that the streamflow response to a long-term change can be understood from interannual variability. This study confirms that the observed shifts in catchment behaviour during the Millennium drought are more related to catchment biophysical structure than exogenous factors such as higher temperatures during the drought. We found that catchments with higher variability in groundwater storage are more likely to show different response to short (~single year) and long (~multiyear) droughts. Larger shifts in the rainfall-runoff relationship (i.e. more susceptible to change behaviour if the climate becomes drier) are also more likely in arid catchments, catchments with less variable monthly rainfall (i.e. less seasonality), and catchments with deeper soils. We found that spring rainfall deficits, even though small, were still important. However, the importance of the spring rainfall deficits is likely to be specific for the study region and possibly the dry period investigated in this study. Clear agreement between the proportion of evidence results obtained with notably different performance criteria is reassuring, suggesting that the results are robust. Our results are pertinent for understanding changes in watershed behaviour associated with changes in regional climate.

Acknowledgements

We thank Anne Van Loon and two anonymous reviewers for comments and suggestions that improved the paper. Murray Peel is the recipient of an Australian Research Council Future Fellowship (FT120100130). We thank Jai Vaze for providing streamflow and catchment 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 state government agencies online.

References

Akaike, H. (1973), Information theory and an extension of the maximum likelihood principle. paper presented at Second International Symposium on Information Theory, Akademinai Kiado.

Anderson, D. R., K. P. Burnham, and G. C. White (1998), Comparison of Akaike information criterion and consistent Akaike information criterion for model selection and statistical inference from capture-recapture studies, *Journal of Applied Statistics*, *25*(2), 263-282.

Angers, D. A., and J. Caron (1998), Plant-induced changes in soil structure: Processes and feedbacks, *Biogeochemistry*, *42*(1-2), 55-72.

Berghuijs, W. R., M. Sivapalan, R. A. Woods, and H. H. G. Savenije (2014), Patterns of similarity of seasonal water balances: A window into streamflow variability over a range of time scales, *Water Resour. Res.*, *50*(7), 5638-5661.

Beven, K. (2010), *Environmental modelling: An uncertain future?*, CRC Press.

Bozdogan, H. (1987), Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions, *Psychometrika*, *52*(3), 345-370.

Brunner, P., C. T. Simmons, and P. G. Cook (2009), Spatial and temporal aspects of the transition from connection to disconnection between rivers, lakes and groundwater, *J. Hydrol.*, *376*(1-2), 159-169.

Brunner, P., P. G. Cook, and C. T. Simmons (2011), Disconnected surface water and groundwater: From theory to practice, *Ground Water*, *49*(4), 460-467.

Brutsaert, W. (2008), Long-term groundwater storage trends estimated from streamflow records: Climatic perspective, *Water Resour. Res.*, *44*(2), W02409.

Budyko, M. I. (1974), *Climate and Life*, 508 pp., Academic, San Diego, California.

Burnham, K. P., and D. R. Anderson (2002), *Model selection and multimodel inference : a practical information-theoretic approach*, 2nd ed. ed., Springer, New York.

Cai, W., and T. Cowan (2008), Evidence of impacts from rising temperature on inflows to the Murray-Darling Basin, *Geophysical Research Letters*, *35*(7), doi: 10.1029/2008gl033390.

Chen, J. L., C. R. Wilson, B. D. Tapley, B. Scanlon, and A. Güntner (2016), Long-term groundwater storage change in Victoria, Australia from satellite gravity and in situ observations, *Global and Planetary Change*, *139*, 56-65.

Chiew, F. H. S., M. C. Peel, T. A. McMahon, and L. W. Siriwardena (2006), Precipitation elasticity of streamflow in catchments across the world, paper presented at FRIEND 2006, IAHS, Havana, Cuba.

Chiew, F. H. S., W. J. Young, W. Cai, and J. Teng (2011), Current drought and future hydroclimate projections in southeast Australia and implications for water resources management, *Stoch. Environ. Res. Risk Assess.*, *25*(4), 601-612.

Chiew, F. H. S., A. J. Frost, R. Srikanthan, T. A. McMahon, M. Peel, J. M. Austin, and N. Potter (2008), *Characterisation of Recent Rainfall and Runoff in the Murray-Darling Basin: A Report to the Australian Government from the CSIRO Murray-Darling Basin Sustainable Yields Project*, CSIRO.

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, *Stoch. 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*(5), W05552, doi:05510.01029/02011WR011721.

CSIRO (2012), Climate and water availability in south-eastern Australia: A synthesis of findings from Phase 2 of the South Eastern Australian Climate Initiative (SEACI)*Rep.*, 41 pp, CSIRO, Australia.
d'Agostino, R. B. (1971), An omnibus test of normality for moderate and large size samples, *Biometrika*, *58*(2), 341-348.
d'Agostino, R. B., and E. S. Pearson (1973), Tests for departure from normality. Empirical results for the distributions of b2 and√ b1, *Biometrika*, *60*(3), 613-622.
Dunne, T., and R. D. Black (1970), Partial area contributions to storm runoff in a small New England watershed, *Water Resour. Res.*, *6*(5), 1296-1311.
Eltahir, E. A. B., and P. J. F. Yeh (1999), On the asymmetric response of aquifer water level to floods and droughts in Illinois, *Water Resour. Res.*, *35*(4), 1199-1217.
Ezekiel, M. (1929), The application of the theory of error to multiple and curvilinear correlation, *Journal of the American Statistical Association*, *24*(165A), 99-104.
Farmer, D., M. Sivapalan, and C. Jothityangkoon (2003), Climate, soil, and vegetation controls upon the variability of water balance in temperate and semiarid landscapes: Downward approach to water balance analysis, *Water Resour. Res.*, *39*(2), SWC11-SWC121.
Gao, H., M. Hrachowitz, S. J. Schymanski, F. Fenicia, N. Sriwongsitanon, and H. H. G. Savenije (2014), Climate controls how ecosystems size the root zone storage capacity at catchment scale, *Geophysical Research Letters*, *41*(22), 7916-7923.
Gudmundsson, L., P. Greve, and S. I. Seneviratne (2016), The sensitivity of water availability to changes in the aridity index and other factors—A probabilistic analysis in the Budyko space, *Geophysical Research Letters*, *43*(13), 6985-6994.
Hisdal, H., K. Stahl, L. M. Tallaksen, and S. Demuth (2001), Have streamflow droughts in Europe become more severe or frequent?, *International Journal of Climatology*, *21*(3), 317-333.
Hughes, J. D., K. C. Petrone, and R. P. Silberstein (2012), Drought, groundwater storage and stream flow decline in southwestern Australia, *Geophysical Research Letters*, *39*(3), L03408.
Hurvich, C. M., and C.-L. Tsai (1989), Regression and time series model selection in small samples, *Biometrika*, *76*(2), 297-307.
Jeffrey, S. J., J. O. Carter, K. B. Moodie, and A. R. Beswick (2001), Using spatial interpolation to construct a comprehensive archive of Australian climate data, *Environmental Modelling and Software*, *16*(4), 309-330.
Kinal, J., and G. L. Stoneman (2012), Disconnection of groundwater from surface water causes a fundamental change in hydrology in a forested catchment in south-western Australia, *J. Hydrol.*, *472-473*, 14-24.
Leblanc, M. J., P. Tregoning, G. Ramillien, S. O. Tweed, and A. Fakes (2009), Basin-scale, integrated observations of the early 21st century multiyear drought in Southeast Australia, *Water Resour. Res.*, doi:10.1029/2008WR007333.
Loeb, N. G., H. Wang, L. Liang, S. Kato, and F. G. Rose (2016), Surface energy budget changes over Central Australia during the early 21st century drought, *International Journal of Climatology*, doi:10.1002/joc.4694.
Merz, R., J. Parajka, and G. Blöschl (2011), Time stability of catchment model parameters: Implications for climate impact analyses, *Water Resour. Res.*, *47*(2), W02531.
Montanari, A., et al. (2013), "Panta Rhei-Everything Flows": Change in hydrology and society-The IAHS Scientific Decade 2013-2022, *Hydrological Sciences Journal*, *58*(6), 1256-1275.
Montgomery, D. C., and E. A. Peck (1992), *Introduction to linear regression analysis*, 2 ed., John Wiley & Sons.

Mueller, R. C., C. M. Scudder, M. E. Porter, R. Talbot Trotter, C. A. Gehring, and T. G. Whitham (2005), Differential tree mortality in response to severe drought: evidence for long-term vegetation shifts, *Journal of Ecology*, *93*(6), 1085-1093.

Murphy, B. F., and B. Timbal (2008), A review of recent climate variability and climate change in Southeastern Australia, *International Journal of Climatology*, *28*(7), 859-879.

Peel, M. C., and G. Blöschl (2011), Hydrological modelling in a changing world, *Progress in Physical Geography*, *35*(2), 249-261.

Peters, E., P. J. J. F. Torfs, H. A. J. van Lanen, and G. Bier (2003), Propagation of drought through groundwater - A new approach using linear reservoir theory, *Hydrological Processes*, *17*(15), 3023-3040.

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, pp. 3622-3628, Perth, WA.

Petrone, K. C., J. D. Hughes, T. G. Van Niel, and R. P. Silberstein (2010), Streamflow decline in southwestern Australia, 1950-2008, *Geophysical Research Letters*, *37*(11), 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*(10), W00G10.

Potter, N. J., F. H. S. Chiew, and A. J. Frost (2010), An assessment of the severity of recent reductions in rainfall and runoff in the Murray-Darling Basin, *J. Hydrol.*, *381*(1-2), 52-64.

Potter, N. J., C. Petheram, and L. Zhang (2011), Sensitivity of streamflow to rainfall and temperature in south-eastern Australia during the Millennium drought, in *MODSIM 2011 - 19th International Congress on Modelling and Simulation - Sustaining Our Future: Understanding and Living with Uncertainty*, edited, pp. 3636-3642, Perth, WA.

Roderick, M. L., and G. D. Farquhar (2011), A simple framework for relating variations in runoff to variations in climatic conditions and catchment properties, *Water Resour. Res.*, *47*(6), doi: 10.1029/2010wr009826.

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.

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, *Geophysical Research Letters*, *43*(4), 1574-1581.

Stephens, M. A. (1974), EDF statistics for goodness of fit and some comparisons, *Journal of the American statistical Association*, *69*(347), 730-737.

Stoelzle, M., K. Stahl, A. Morhard, and M. Weiler (2014), Streamflow sensitivity to drought scenarios in catchments with different geology, *Geophysical Research Letters*, *41*(17), 6174-6183.

Tallaksen, L. M., and H. A. J. Van Lanen (2004), *Hydrological drought: processes and estimation methods for streamflow and groundwater*, Elsevier.

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*(7), 4903-4922.

Tromp-Van Meerveld, H. J., and J. J. McDonnell (2006), Threshold relations in subsurface stormflow: 2. The fill and spill hypothesis, *Water Resour. Res.*, *42*(2).

UNEP (1992), *World atlas of desertification / United Nations Environment Programme*, Edward Arnold, London ; Baltimore.

Van Lanen, H. A. J., N. Wanders, L. M. Tallaksen, and A. F. Van Loon (2013), Hydrological drought across the world: Impact of climate and physical catchment structure, *Hydrology and Earth System Sciences*, *17*(5), 1715-1732.

Van Loon, A. F. (2015), Hydrological drought explained, *Wiley Interdisciplinary Reviews: Water*, *2*(4), 359-392.

Van Loon, A. F., and G. Laaha (2015), Hydrological drought severity explained by climate and catchment characteristics, *J. Hydrol.*, *526*, 3-14.

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

Vaze, J., F. H. S. Chiew, J. M. Perraud, N. Viney, D. Post, J. Teng, B. Wang, J. Lerat, and M. Goswami (2010b), Rainfall-runoff modelling across southeast Australia: Datasets, models and results, *Australian Journal of Water Resources*, *14*(2), 101-116.

Verdon-Kidd, D. C., and A. S. Kiem (2009), Nature and causes of protracted droughts in southeast Australia: Comparison between the Federation, WWII, and Big Dry droughts, *Geophys. Res. Lett.*, *36*(22), L22707.

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*(5).

Wang, Z., and B. Thompson (2007), Is the Pearson r 2 biased, and if so, what is the best correction formula?, *The Journal of Experimental Education*, *75*(2), 109-125.

Western, A. W., G. Blöschl, and R. B. Grayson (2001), Toward capturing hydrologically significant connectivity in spatial patterns, *Water Resour. Res.*, *37*(1), 83-97.

Western, A. W., R. B. Grayson, and G. Blöschl (2002), Scaling of soil moisture: A hydrologic perspective, *Annual Review of Earth & Planetary Science*, *30*, 149-180.

Figure 1.

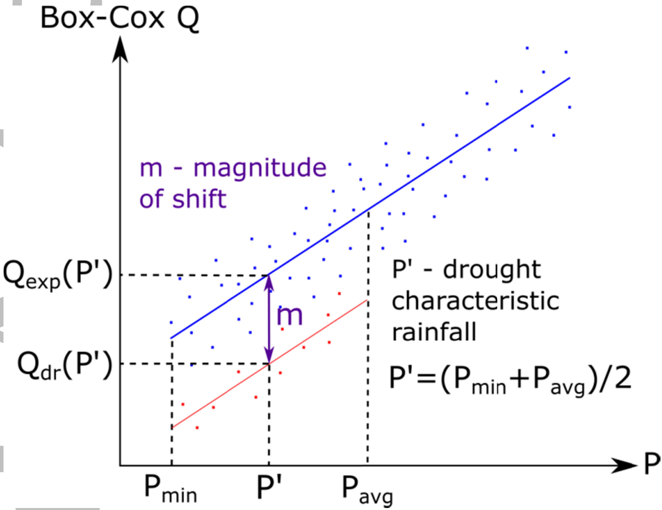

Box-Cox Q
m - magnitude
of shift
$Q_{exp}(P')$
$Q_{dr}(P')$
m
P' - drought
characteristic
rainfall
$P'=(P_{min}+P_{avg})/2$
P
P_{min}
P'
P_{avg}

Figure 2.

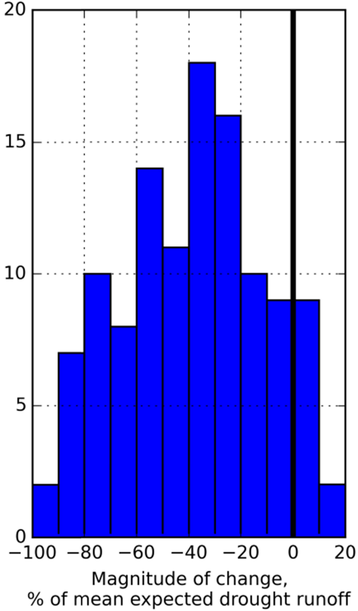

20
15
10
5
0
−100
−80
−60
−40
−20
0
20
Magnitude of change,
% of mean expected drought runoff

Figure 3.

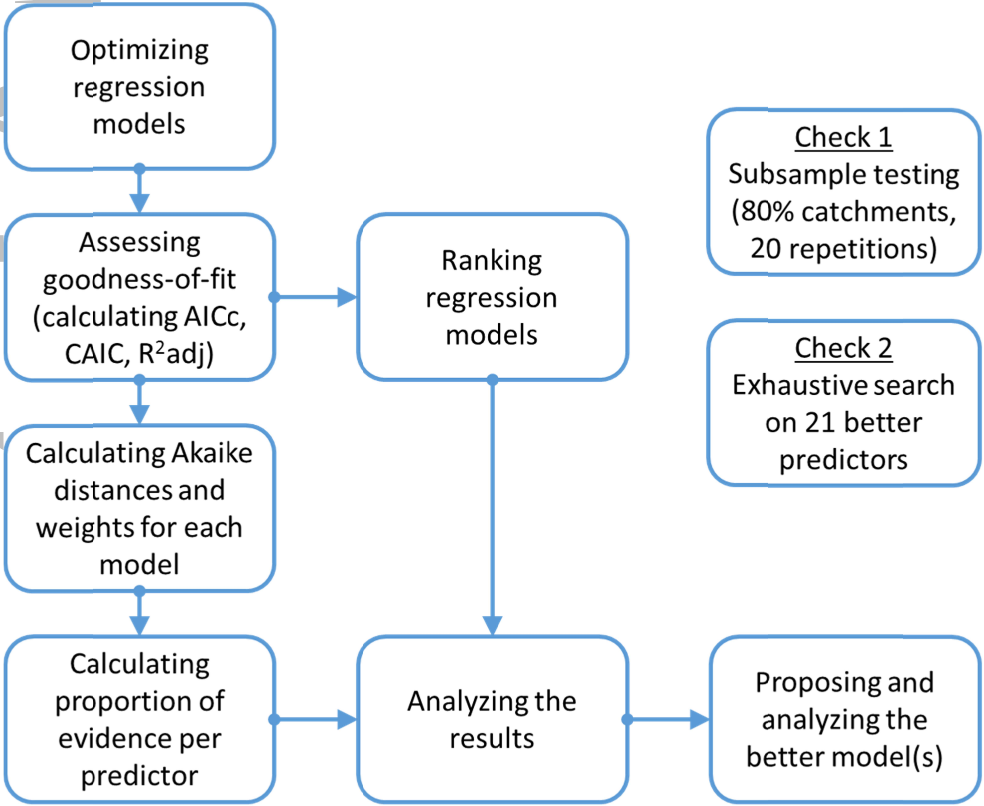

Optimizing regression models
Assessing goodness-of-fit (calculating AICc, CAIC, R^2adj)
Calculating Akaike distances and weights for each model
Calculating proportion of evidence per predictor
Ranking regression models
Analyzing the results
Check 1
Subsample testing (80% catchments, 20 repetitions)
Check 2
Exhaustive search on 21 better predictors
Proposing and analyzing the better model(s)

Figure 4.

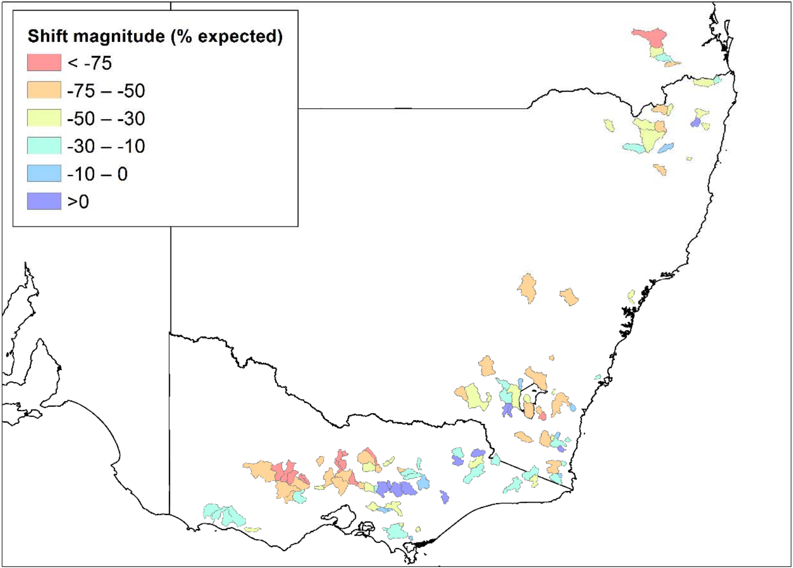

Shift magnitude (% expected)
< -75
-75 – -50
-50 – -30
-30 – -10
-10 – 0
>0

Figure 5.

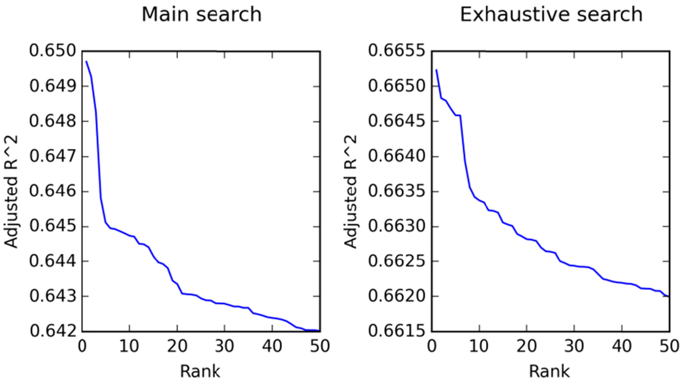

Main search
Adjusted R^2
0.650
0.649
0.648
0.647
0.646
0.645
0.644
0.643
0.642
0
10
20
30
40
50
Rank
Exhaustive search
Adjusted R^2
0.6655
0.6650
0.6645
0.6640
0.6635
0.6630
0.6625
0.6620
0.6615
0
10
20
30
40
50
Rank

Figure 6.

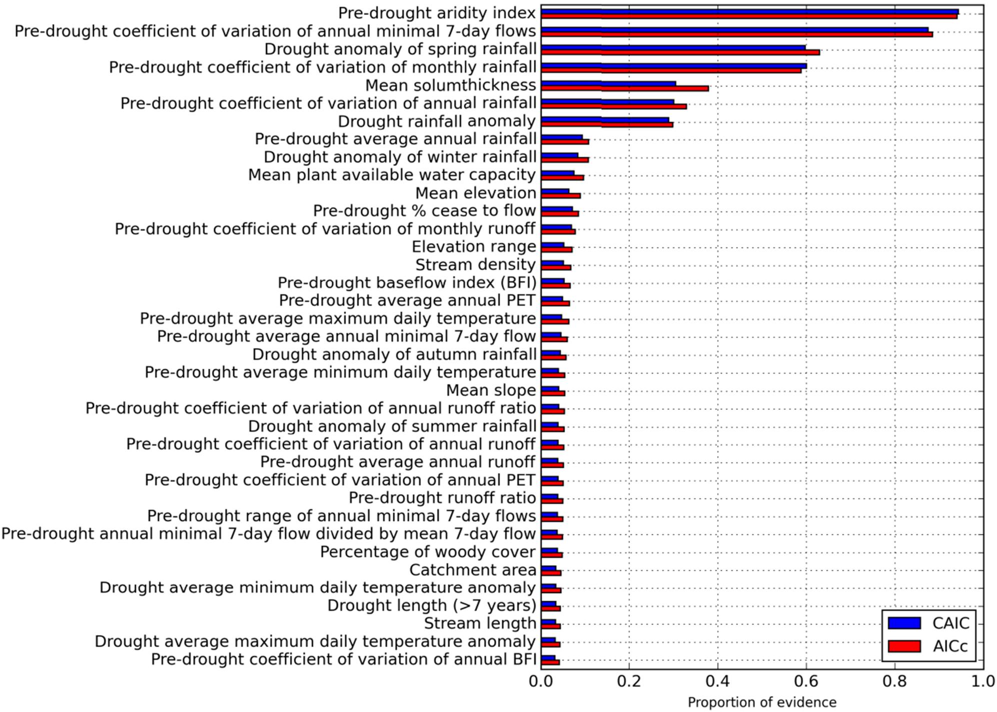

Pre-drought aridity index
Pre-drought coefficient of variation of annual minimal 7-day flows
Drought anomaly of spring rainfall
Pre-drought coefficient of variation of monthly rainfall
Mean solumthickness
Pre-drought coefficient of variation of annual rainfall
Drought rainfall anomaly
Pre-drought average annual rainfall
Drought anomaly of winter rainfall
Mean plant available water capacity
Mean elevation
Pre-drought % cease to flow
Pre-drought coefficient of variation of monthly runoff
Elevation range
Stream density
Pre-drought baseflow index (BFI)
Pre-drought average annual PET
Pre-drought average maximum daily temperature
Pre-drought average annual minimal 7-day flow
Drought anomaly of autumn rainfall
Pre-drought average minimum daily temperature
Mean slope
Pre-drought coefficient of variation of annual runoff ratio
Drought anomaly of summer rainfall
Pre-drought coefficient of variation of annual runoff
Pre-drought average annual runoff
Pre-drought coefficient of variation of annual PET
Pre-drought runoff ratio
Pre-drought range of annual minimal 7-day flows
Pre-drought annual minimal 7-day flow divided by mean 7-day flow
Percentage of woody cover
Catchment area
Drought average minimum daily temperature anomaly
Drought length (>7 years)
Stream length
Drought average maximum daily temperature anomaly
Pre-drought coefficient of variation of annual BFI
CAIC
AICc
0.0
0.2
0.4
0.6
0.8
1.0
Proportion of evidence

Figure 7.

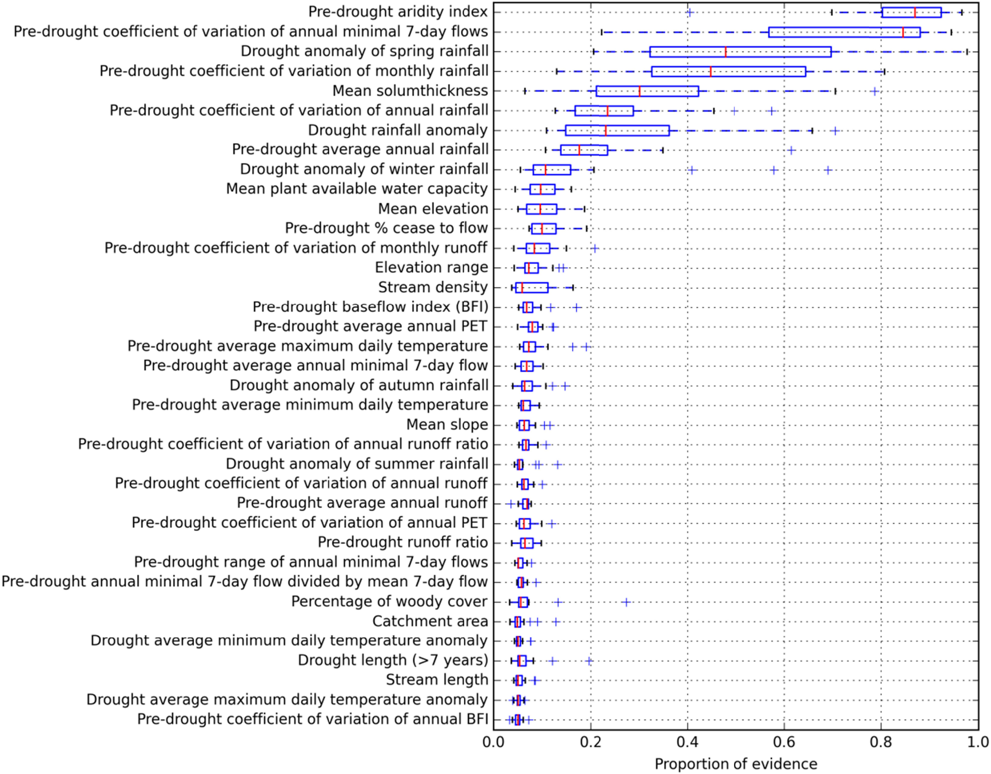
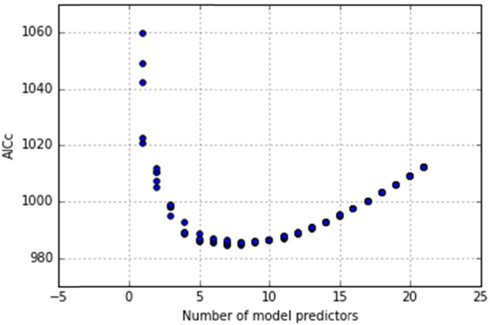
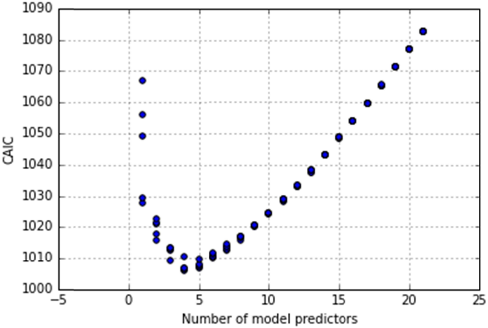
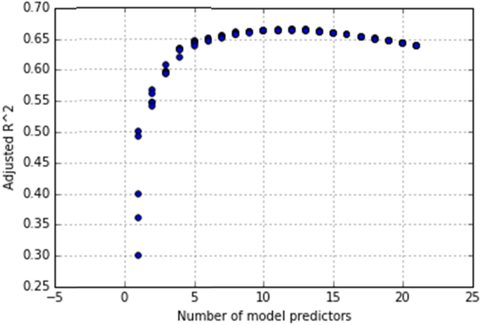

Pre-drought aridity index
Pre-drought coefficient of variation of annual minimal 7-day flows
Drought anomaly of spring rainfall
Pre-drought coefficient of variation of monthly rainfall
Mean solumthickness
Pre-drought coefficient of variation of annual rainfall
Drought rainfall anomaly
Pre-drought average annual rainfall
Drought anomaly of winter rainfall
Mean plant available water capacity
Mean elevation
Pre-drought % cease to flow
Pre-drought coefficient of variation of monthly runoff
Elevation range
Stream density
Pre-drought baseflow index (BFI)
Pre-drought average annual PET
Pre-drought average maximum daily temperature
Pre-drought average annual minimal 7-day flow
Drought anomaly of autumn rainfall
Pre-drought average minimum daily temperature
Mean slope
Pre-drought coefficient of variation of annual runoff ratio
Drought anomaly of summer rainfall
Pre-drought coefficient of variation of annual runoff
Pre-drought average annual runoff
Pre-drought coefficient of variation of annual PET
Pre-drought runoff ratio
Pre-drought range of annual minimal 7-day flows
Pre-drought annual minimal 7-day flow divided by mean 7-day flow
Percentage of woody cover
Catchment area
Drought average minimum daily temperature anomaly
Drought length (>7 years)
Stream length
Drought average maximum daily temperature anomaly
Pre-drought coefficient of variation of annual BFI
0.0
0.2
0.4
0.6
0.8
1.0
Proportion of evidence

Figure 8.


AICc
1060
1040
1020
1000
980
−5
0
5
10
15
20
25
Number of model predictors


CAIC
1090
1080
1070
1060
1050
1040
1030
1020
1010
1000
−5
0
5
10
15
20
25
Number of model predictors


Adjusted R^2
0.70
0.65
0.60
0.55
0.50
0.45
0.40
0.35
0.30
0.25
−5
0
5
10
15
20
25
Number of model predictors