

Long-range forecasting of intermittent streamflow

F. F. van Ogtrop¹, R. W. Vervoort¹, G. Z. Heller², D. M. Stasinopoulos³, and R. A. Rigby³

¹Hydrology Research Laboratory Faculty of Agriculture, Food and Natural Resources, the University of Sydney, NSW, Sydney, Australia

²Department of Statistics, Macquarie University, NSW, Sydney, Australia

³Statistics, OR, and Mathematics (STORM) Research Centre, London Metropolitan University, London, UK

Received: 13 December 2010 – Published in Hydrol. Earth Syst. Sci. Discuss.: 20 January 2011

Revised: 27 August 2011 – Accepted: 13 October 2011 – Published: 7 November 2011

Abstract. Long-range forecasting of intermittent streamflow in semi-arid Australia poses a number of major challenges. One of the challenges relates to modelling zero, skewed, non-stationary, and non-linear data. To address this, a statistical model to forecast streamflow up to 12 months ahead is applied to five semi-arid catchments in South Western Queensland. The model uses logistic regression through Generalised Additive Models for Location, Scale and Shape (GAMLSS) to determine the probability of flow occurring in any of the systems. We then use the same regression framework in combination with a right-skewed distribution, the Box-Cox t distribution, to model the intensity (depth) of the non-zero streamflows. Time, seasonality and climate indices, describing the Pacific and Indian Ocean sea surface temperatures, are tested as covariates in the GAMLSS model to make probabilistic 6 and 12-month forecasts of the occurrence and intensity of streamflow. The output reveals that in the study region the occurrence and variability of flow is driven by sea surface temperatures and therefore forecasts can be made with some skill.

1 Introduction

Predictions of rainfall and river flows over long time scales can provide many benefits to agricultural producers (Abawi et al., 2005; Brown et al., 1986; Mjelde et al., 1988; Wilks and Murphy, 1986; White, 2000). Predicting these variables in semi-arid regions is especially difficult because of extreme spatial and temporal variability of both climate and streamflow (Chiew et al., 2003). In addition, data are often scarce, possibly due to many semi-arid regions supporting low human populations. Previous models to predict rainfall and

streamflow in semi-arid areas have had low accuracy, which has led to criticism by farmers, who are the key users of this information (Hayman et al., 2007). The challenge is thus to develop accurate forecasts for highly variable systems with minimal data requirements.

Forecasting streamflow in semi-arid regions poses a number of further hurdles. A model of semi-arid streamflow needs to be able to cope with extensive zeroes, extremely skewed, locally non-stationary, and non-linear data (Yakovitz, 1973; Milly et al., 2008). However, on a positive note, modelling data with a positive density at zero can be achieved by dealing with the zero and non-zero data separately. Examples of such two-part models can be found in the modelling of species abundance (Barry and Welsh, 2002), rainfall (Hyndman and Grunwald, 2000), medicine (Lachenbruch, 2001) and insurance claims (De Jong and Heller, 2008). Furthermore, generalised additive models (GAM) can model non-normal (skewed) data and non-linear relationships between the streamflow and potential predictors (Hastie and Tibshirani, 1986; Wood, 2006). Trends, or non-stationarity, in the data can be accounted for by adding synthetic variables as covariates in such models (Hyndman and Grunwald, 2000; Heller et al., 2009; Grunwald and Jones, 2000).

Forecasting streamflow directly from climate indices has shown promise, as the relation between streamflow and climate tends to be stronger than for rainfall (Wooldridge et al., 2001). One of the key climatological parameters driving streamflow throughout Australia is the El Niño Southern Oscillation (ENSO) which describes variations in sea surface temperatures (SST) in the Pacific Ocean (Chiew et al., 1998, 2003; Dettinger and Diaz, 2000; Dutta et al., 2006; Piechota et al., 1998). More recently, effects of the Indian Ocean SST on South Eastern Australian rainfall have been suggested (Cai et al., 2009; Ummenhofer et al., 2009; Verdon and Franks, 2005a,b), and recent research suggests that the Indian Ocean is an important driver of streamflow in Victoria,



Correspondence to: F. F. van Ogtrop
(floris.vanogtrop@sydney.edu.au)

Australia (Kiem and Verdon-Kidd, 2009). As a result, both the Pacific and Indian Ocean sea surface temperatures are considered essential in understanding the full variability of weather patterns and streamflow associated with each ENSO phase (Wang and Hendon, 2007; Kiem and Verdon-Kidd, 2009).

In the past, several researchers have used data-driven approaches to model the relationship between either rainfall or streamflow, and climate indices at various time scales and lags (Table 1). There have been few comparative studies of the techniques listed in Table 1. However, the performance of Generalised Additive Modeling (GAM) compared favourably with Neural Networks (NN) for modelling precipitation (Guisan et al., 2002). Furthermore, in contrast to NN, GAM allows identification of the influence of the individual covariates, which assists in comprehending the underlying physical processes being modelled (Schwarzer et al., 2000; Faraway and Chatfield, 1998). Similarly, GAM has been shown to outperform discriminant analysis (Berg, 2007) which has been used previously to model climate streamflow relationships (Piechota et al., 2001; Piechota and Dracup, 1999). Generalised models for location scale and shape (GAMLSS) (Rigby and Stasinopoulos, 2005) potentially perform better than GAM because a broader selection of distributions is available, which can capture the skewness of streamflow data in semi-arid regions (Heller et al., 2009).

Aside from studies by Sharma et al. (2000) in a more coastal environment and our preliminary study, Heller et al. (2009), there appear to be no other studies that apply GAM or GAMLSS to explore relationships between climate indices and streamflow.

The aim of this study therefore is to test the general ability of GAMLSS to produce 6 and 12 month ahead monthly streamflow forecasts in several large semi-arid river systems. An advantage is that the results can be expressed as a cumulative distribution function, which gives the probability of exceeding threshold flow volumes. This is also known as the flow duration curve. Furthermore, the model uncertainty is intrinsically incorporated in the probabilistic output (Krzysztofowicz, 1983; Jolliffe and Stephenson, 2003; Buizza, 2008; Pappenberger and Beven, 2006; Hamill and Wilks, 1995). Finally, a statistical approach is more suitable for modelling streamflow in these regions, as limited biophysical data and understanding would thwart the use of a more mechanistic modelling approach.

2 Data and methods

2.1 Data

This study considers five river systems in south-western Queensland (SWQ), Australia (Table 2, Fig. 1). All of the river systems are similar, being terminal inland river systems and intermittent in nature. Roughly, the average annual

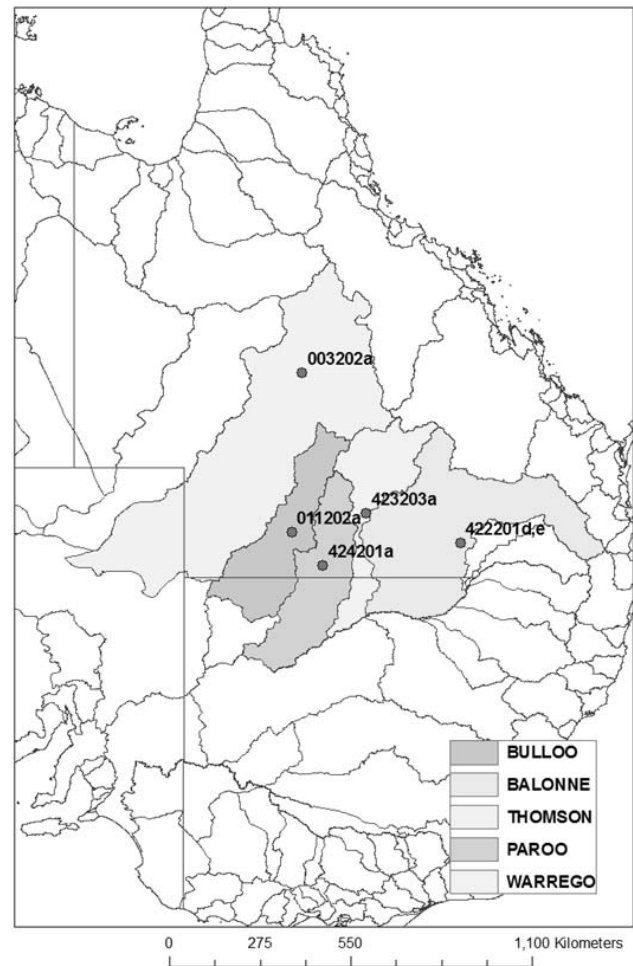


Fig. 1. Location of river gauging stations and surrounding basins, South Western Queensland.

rainfall decreases in a south westerly direction. With the exception of the Balonne, all of the river systems are unregulated. Streamflow in the Balonne River has been altered as a result of water extraction (Thoms and Parsons, 2003; Thoms, 2003) with most of the change occurring in flows with an average occurrence interval of less than 2 years (Thoms, 2003). Hence, an unpaired dataset for this river was also used, which was provided by the Department of Environment and Resources Management in Queensland Australia and was created using the Integrated Quality Quantity Model (IQQM) (Hameed and Podger, 2001; Simons et al., 1996). There is no doubt that the validity of this model and the results might be questioned. However, due to the often controversial nature of streamflow data, this data is the only accessible data which accounts for water extractions in the region. A double mass analysis on the modelled versus measured data indicates an almost perfect agreement between the periods of 1922 and 1950. Thereafter, and coinciding with post second world war urban and rural development, the measured streamflow

Table 1. Summary of statistical models used for forecasting rainfall and streamflow.

Model type	Region	Rainfall/ Streamflow	Time scale	Max Lag	Indices	Author
Artificial Neural Networks	California, USA	Rainfall	Annual	1	ENSO, 700-mb height anomaly	Silverman and Dracup (2000)
Self Organising Linear Output map	Murray Darling Basin	Rainfall	Monthly	12	ENSO	Barros and Bowden (2008)
Linear Correlation/continuous exceedance probability curve	Suwanee River USA	Streamflow	Monthly	9	ENSO	Tootle and Piechota (2004)
Linear Correlation/continuous exceedance probability curve	North Platte River USA	Streamflow	Monthly	6	SST, 500 mb height anomaly	Soukup et al. (2009)
Linear regression	Australia	Streamflow	Monthly	6–12	Indo-Pacific Thermocline	Ruiz et al. (2007)
Linear discriminant analysis	Columbia River Basin, USA	Streamflow	Monthly	7	ENSO	Piechota and Dracup (1999)
Generalised Additive Models	Melbourne, Australia	Rainfall	Daily	0	Only SOI	Hyndman and Grunwald (2000)
Generalized Additive Models	Mauritius	Rainfall	Daily	0	None	Underwood (2009)
Generalised Additive Models	Warragamba Dam, NSW, Australia	Streamflow	Monthly	15	ENSO	Sharma et al. (2000)
Generalised Additive Model	Balonne River, QLD, Australia	Streamflow	Monthly	0	ENSO	Heller et al. (2009)
Nonparametric Kernel	Warragamba dam, Sydney, Australia	Rainfall	Seasonal	6	ENSO	Sharma (2000)
Bayesian joint probability	Murrumbidgee River, Australia	Streamflow	Seasonal	2	ENSO	Wang et al. (2009)
Categorical composites	Williams River, NSW, Australia	Streamflow	Monthly	9	ENSO	Kiem and Franks (2001)
Partitioning	Eastern Australia	Rainfall	Seasonal	1	SOI, GpH*	Cordery (1999)

* GpH = Geopotential Height

Table 2. Flow statistics for south western Queensland Rivers.

River	Station number	Approx. total catchment area km ²	Median m ³ s ⁻¹	Mean flow m ³ s ⁻¹	Standard deviation m ³ s ⁻¹	Coef. of variation σ/μ	% Cease flow
Thomson	003202a	266 469	0.02	40.47	208.49	5.15	47
Bulloo	011202a	69 244	1.4	22.8	78.5	3.45	16
Paroo	424201a	68 589	0.80	16.20	52.30	3.23	27
Warrego	423203a	57 176	0.26	16.99	74.87	4.41	33
Balonne	422201d,e	148 777	1.85	34.50	109.59	3.18	12
Balonne Naturalised	NA	148 777	3.76	46.88	134.68	2.87	6

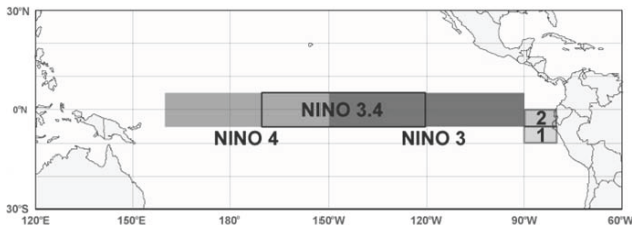


Fig. 2. Locations of average sea surface temperature locations for Niño 1, 2, 3, 3.4 and 4 (source: Bureau of Meteorology, Australia).

decreases relative to the modelled streamflow. Throughout this study, streamflow is given as cubic meters per second (m³ s⁻¹).

Sea surface temperature (SST) data can be readily obtained from several organisations (Table 3). These datasets are usually a combination of spatially averaged monthly temperature in degrees Celsius for various regions of the ocean (Fig. 2) (Wang et al., 1999). For ease of reading the regression formulas, Niño1 + 2 is referred to as Niño1.2. The IOD is the difference between SST in the western and eastern equatorial Indian Ocean (Fig. 3).

Climate datasets prior to 1959 were not considered due to recognised poor data quality (Saji and Yamagata, 2003). Furthermore, the time span of the monthly dataset was reduced to the years 1970 to 2005, which is the maximum length of the monthly flow records for the Bulloo River.

2.2 Models

Modelling zero and non-zero data separately is equivalent to modelling streamflow using a zero-adjusted distribution of the type:

$$f(y; \theta, \pi) = \begin{cases} (1 - \pi) & \text{if } y = 0 \\ \pi f_T(y, \theta) & \text{if } y > 0 \end{cases} \quad (1)$$

where π is the probability of the occurrence of non-zero flow and $f_T(y, \theta)$ is the distribution of the non-zero flow. Hence, initially the occurrence of monthly flow was modelled, for which the results are discussed in Sect. 3.1. As the outcome

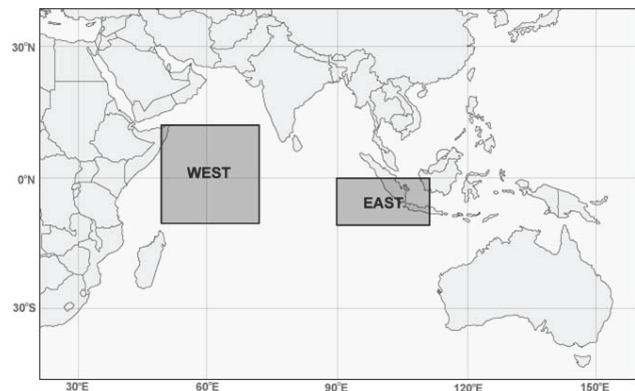


Fig. 3. Locations of average sea surface temperature locations for IOD (source: Bureau of Meteorology, Australia).

is binary, a binomial distribution was used (Hyndman and Grunwald, 2000). As a second step, the intensities (volumes) of the non-zero flows are modelled and the results are discussed in Sect. 3.2.

For the binomial model of the occurrence of flow, the following generalized linear model (GLM) can be initially specified

$$g(\pi) = \log\left(\frac{\pi}{1 - \pi}\right) = \mathbf{x}'\boldsymbol{\beta} \quad (2)$$

where π is the probability of occurrence of non-zero flow, \mathbf{x}' is a vector of covariates, $g(\pi)$ is the logit link function and $\boldsymbol{\beta}$ is a vector of coefficients for x . For comparison, the following GAMLSS was specified (because GAMLSS is an extension of GLM; Rigby and Stasinopoulos, 2001):

$$g(\pi) = \log\left(\frac{\pi}{1 - \pi}\right) = \mathbf{x}'\boldsymbol{\beta} + \sum_{j=1}^J s_j(w_j) \quad (3)$$

where $\mathbf{x}'\boldsymbol{\beta}$ is a combination of linear estimators as in Eq. (2), w_j for $j = 1, 2, \dots, J$ are covariates and s_j for $j = 1, 2, \dots, J$ are smoothing (spline) terms. The addition of smoothing terms in GAMLSS has many advantages, such as identifying non-linear covariate effects in otherwise noisy data sets

Table 3. Summary of data used and availability.

Index	Description	Source	References
Streamflow	Monthly Streamflow (ML/month)	Department of Natural Resources and Water, Queensland http://www.derm.qld.gov.au/water/monitoring/current_data/map_qld.php	
Niño1+2, Niño3, Niño3.4, Niño4	Niño: Averaged Eastern, Central and Western Pacific SST	National Oceanic & Atmospheric Administration, USA http://www.cpc.ncep.noaa.gov/data/indices/sstoi.indices	Trenberth and Stepaniak (2001); Wang et al. (1999)
IOD	Relationship between SST in the eastern equatorial and western equatorial Indian Ocean. Derived from HadISST dataset	Frontier Research Centre for Global Change, Japan http://www.jamstec.go.jp/frsgc/research/d1/iod/	Ummenhofer et al. (2009); Cai et al. (2009)

(Hastie and Tibshirani, 1986). In this study the smoothing is based on penalised B-splines which have been shown to be robust to boundary effects common to other smoothing methods (Eilers and Marx, 1996). The degree of smoothing is selected automatically using penalized maximum likelihood in the `gamlss` package (Rigby and Stasinopoulos, 2005). The GAMLSS models were implemented using the `gamlss` function in the `gamlss` package within the open source program R (R Development Core Team, 2011; Rigby and Stasinopoulos, 2005).

For the intensity model, streamflow data was subset to non-zero flow values. The Box-Cox t distribution (BCT) was used to model non-zero streamflow. This four-parameter flexible distribution (Rigby and Stasinopoulos, 2006), has been shown to be a good fit for non-zero flow data from the Balonne River (Heller et al., 2009) and a number of gauging stations located west of the Australian Capital Territory (Wang et al., 2009). In the BCT distribution $\hat{\mu}$ is the median, $\hat{\sigma}$ is the scale parameter (approximately the coefficient of variation), $\hat{\nu}$ is the skewness and $\hat{\tau}$ is the kurtosis of the non-zero flows. The probability for flows above a flow threshold c can be subsequently calculated as:

$$\hat{p}(\text{flow}_i > c) = \hat{\pi}_i p(Z > z_i) \quad (4)$$

where $z_i = \frac{1}{\hat{\sigma}_i \hat{\nu}} \left[\left(\frac{c}{\hat{\mu}_i} \right)^{\hat{\nu}} - 1 \right]$, if $\hat{\nu} \neq 0$ and $Z \sim t_{\hat{\tau}}$ has a t distribution with $\hat{\tau}$ degrees of freedom and where $\hat{\pi}_i$ is the fitted probability of flow occurring in the i th month (Eq. 4) and $\hat{\mu}_i$, $\hat{\sigma}_i$, $\hat{\nu}$ and $\hat{\tau}$ are the parameters of the fitted BCT distribution. The probability (Eq. 4) can be calculated readily in the `gamlss` package as $\hat{\pi}_i [1 - \text{pBCT}(c, \hat{\mu}_i, \hat{\sigma}_i, \hat{\nu}, \hat{\tau})]$ where `pBCT` is the cumulative distribution function for the BCT distribution (Rigby and Stasinopoulos, 2006). The results of the probability of exceeding a flow threshold are discussed in Sect. 3.3.

2.3 Covariates

Because our interest is in a 1 year ahead forecast, this study focuses on the 12-month lagged covariates as predictors (this means forecasts are based on SST 12 months prior). Water users in the regions expressed most interest in a 12-month ahead forecast as this was perceived to be most beneficial for agricultural planning. Different lag times or combinations of different lag times may also be considered, as point of comparison, 6 month ahead forecasts are also considered. Short and medium range forecasts require additional parameters and a modification to the model type and this is a topic of ongoing research.

A synthetic temporal covariate `Time`, a sequence of consecutive numbers 1, ..., n , where n is the length of the dataset can be included to account for known but unmeasurable or unknown non-stationarity in the data. An example of this could be non-stationarity due to water extraction or as a result of climate change in Eastern Australia (McAlpine et al., 2007; Pitman and Perkins, 2008; Cai and Cowan, 2008; Chiew et al., 2009). A Kwiatkowski-Phillips-Schmidt-Shin test for trend stationarity on the five streamflow datasets (Kwiatkowski et al., 1992) revealed that only the raw Balonne river is non-stationary ($p=0.03$). Hence, `Time` was included to account for non-stationarity due to water extractions for this dataset. A problem with covariates such as `Time` in forecasts is that the future relationship between the response variable and the covariate is unknown and that the relationship is strictly empirical. We can only assume that the observed trend in the data continues for the next 12 months to be used in the forecast. However, the same is somewhat true for all relationships in a statistical model, but in contrast, for the SST covariates, we can assume that there is some underlying physical process which is captured by the statistical model. For a slowly varying smooth covariate the lack of knowledge about future trends might also not

be a concern, but for a rapidly changing covariate (or jump changes) it could be problematic.

The synthetic variable sine is a harmonic covariate included to account for seasonal fluctuations in the data (Hyndman and Grunwald, 2000):

$$\text{sine} = \sin\left(\frac{2\pi S_m}{12}\right) \quad (5)$$

where S_m is $m \pmod{12}$ where m is the month. The sine term was included in each model outside the stepwise procedure. This ensured that strongly seasonal nature of these river systems was accounted for in each model and that the relationship between the SST and streamflow is not due to the fact that both datasets are seasonal. Fitting higher order harmonics was not deemed necessary due to the added flexibility of fitting the harmonic covariate with a penalised B-spline. Dominance of this covariate indicates strong seasonality in the streamflow and thus this term captures seasonal climatic or within catchment processes.

In essence the overall structure means we assume a layered catchment scale model to explain the variation in the streamflow. The first layer consists of the within catchment processes and seasonal variations (what would normally be the main focus of catchment hydrology) captured in the harmonic term. The second layer represents the influence of SSTs and thus it is assumed the influx of moisture from oceanic sources. The final layer consists of a long term trends or periodicities such as caused by water extractions, natural cycles or climate change.

In all cases, the models were assumed to be additive (Sharma et al., 2000). Initial explorative testing of incorporating interaction terms in the form of smoothing surfaces using locally weighted scatterplot smoothing (loess) revealed no improvement in the models.

2.4 Goodness of fit

To determine the most parsimonious model (the best model with the least number of covariates), a stepwise fitting method, the stepGAIC function, is used. This is based on the Generalized Akaike Information Criterion (GAIC), which is a model selection criterion where $\text{GAIC} = -2L + kN$, L is the log likelihood, k is the penalty parameter and N is the number of parameters in the fitted model (Akaike, 1974). A value of $k = 2$ was used as this gave good skill in most models selected and retained more of the SST covariates compared to using higher values of k . The stepGAIC process also selects whether or not B-splines are fitted to the covariates. Hence, it is quite possible that the most parsimonious model is simply a GLM. Forward backward selection gave superior results to only backward selection and using the full model. The model residuals were checked for independence and identical distribution.

Validation of the models was conducted using a leave 12 month out cross validation routine (Chowdhury and Sharma,

2009; Wilks, 2005). Essentially, this involved leaving one year (12 months) of data out in each model run and then using the left out data for the final forecast. Forecast skill was then calculated based on the combined forecasts of the cross validated results.

The Brier Skill Score (BSS) and Relative Operating Characteristic (ROC) are the most common means for verifying probabilistic forecasts (Jolliffe and Stephenson, 2003; Wilks, 2006). These were implemented in the verification package in R (NCAR, 2010). The BSS ranges from 0 to 1 where 0 indicates no skill and 1 indicates a perfect forecast and the ROC is presented as a p-value which test the null-hypothesis that there is no forecast skill (Mason and Graham, 2002). Any value less than 0.01 is taken to be significant. Typical BSS values for forecasts of daily streamflow in a temperate climate lie between 0.6 and 0.8 at day one and decrease to between 0 and 0.2 at day 10 (Roulin and Vannitsem, 2005). Similarly, BSS values of between 0 and 0.5 were found in Iowa (USA) using monthly ensemble streamflow prediction (Hashino et al., 2006).

3 Results and discussion

3.1 Occurrence model

Typical examples of the fitted models for the occurrence of non-zero flows for SWQ Rivers are given in Table 4.

The Pacific Ocean SST affects the strength of the northern Australian monsoon and cyclonic activity over a year (Evans and Allan, 1992). Local knowledge suggest that cyclonic activity close to, or crossing, the coast in north eastern Australia is often indicative of significant streamflow in the study region with a delay of up to two months. From Table 4, it is clear that the Pacific Ocean SSTs are drivers of the probability of occurrence of zero streamflow in all of the rivers. The relationship between the eastern Niño1.2 and the central and western Pacific Niño3 and Niño4 are of opposite sign. This may be explained by the fact that changes in SST in the central and western pacific and the eastern Pacific are phase shifted to varying degrees (Wang et al., 2010). Finally, streamflow in the Balonne River, which has one of its two major sources further south east than the other catchments, is significantly affected by IOD. It has been shown that IOD is linked with the development of northwest cloudbands (Verdon and Franks, 2005a) which in turn can bring winter rainfall to central and Eastern Australia (Braganza, 2008; Courtney, 1998; Collins, 1999).

The inclusion of a Time covariate for the raw Balonne River model allows investigation of whether we can account for water extraction occurring upstream of the gauging station. The model indicates that post 1980, the probability of observed flow occurring in the Balonne is decreasing in time (Fig. 4). This would suggest that increased water extraction occurred post 1980 upstream of the gauging station (Thoms,

Table 4. Occurrence models (1 year forecast) for river systems in south western Queensland. In these formulas $\hat{\pi}$ is the fitted probability of occurrence of flow, Time is a sequence 1, 2, 3, ..., n and $s(\cdot)$ is a penalised B-spline smooth function. The other covariates are as described in Table 3.

River	Occurrence model
Thomson	$7.07 + s(\text{sine}) + 0.73 \text{ Niño}1.2 - s(\text{Niño}3) + s(\text{Niño}3.4) - s(\text{Niño}4)$
Bulloo	$6.10 + s(\text{sine}) + 0.35 \text{ Niño}1.2 - s(\text{Niño}4)$
Paroo	$-1.48 + s(\text{sine}) + s(\text{Niño}1.2) - 1.59 \text{ Niño}3 + s(\text{Niño}3.4)$
Warrego	$-1.77 + s(\text{sine}) + 0.70 \text{ Niño}1.2 - 2.09 \text{ Niño}3$
Balonne	$24.80 + s(\text{sine}) + s(\text{Time}) - s(\text{Niño}4) + 0.45 \text{ IOD}$
Balonne Naturalised	$52.27 + s(\text{sine}) + 0.54 \text{ Niño}1.2 - s(\text{Niño}4) + s(\text{IOD})$

2003; Thoms and Parsons, 2003). Rather than using a Time variable it may be possible to include actual extraction volumes or at least a function representing extraction rules as a covariate. The skill scores for the combined model are reported later in Table 6.

3.2 Intensity model

The intensity model gives the probability of the level of non-zero monthly flow above a threshold. It therefore predicts the distribution of monthly flow values (Table 5).

From Table 5, it appears that the entire Pacific has a stronger influence in the 6 month forecast and the eastern and central Pacific in the 12 month forecast. The $\hat{\mu}$ models selected are reasonably homogenous for all rivers. Furthermore, the direction of influence (sign) is consistent for all models. This shows the potential of a stepwise approach for understanding what climate drivers influence which region as suggested by Wang et al. (2009). However, there is some spatial heterogeneity in the relationships between SST and streamflow particularly for the 6 month forecasts and $\hat{\sigma}$ for both the 6 and 12 month forecasts. Furthermore, the covariates selected in the occurrence model (Table 4) are not consistent with those selected for the intensity model. This is the result of the forward backward stepwise covariate selection approach which could select a number of equally plausible models (Whittingham et al., 2006), which makes inference from the output tentative at best. This would give cause for including all parameters in the model rather than using stepwise selection. However, there are also problems associated with using a full model. Importantly for this study is that incorporating non-significant parameters may cause excess noise in the model predictions and thus less skilful forecasts (Whittingham et al., 2006). The trade off between model complexity and skill is a topic for future research. One further important result is that the forecast for the raw Balonne data shows significant skill. This suggests that the Time term adequately accounts for the non-stationarity due

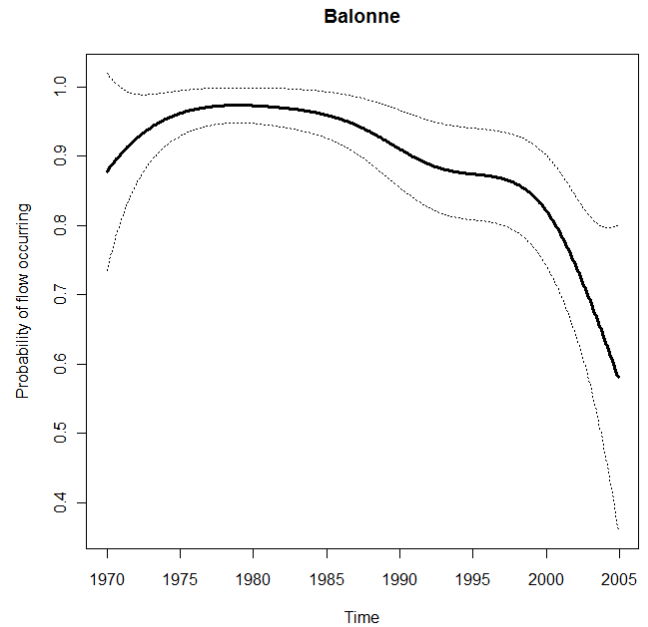


Fig. 4. The fitted B-spline and 95 % confidence intervals (dotted lines) for the Time covariate non-naturalised Balonne river data.

to water extraction and that the naturalisation of the data may not be required. The advantage of this is that uncertainties introduced in the naturalisation process and political sensitivities associated with irrigation water extraction are bypassed.

3.3 Probabilistic forecast of streamflow

Using Eq. (5), the probability of getting at least the median flow was calculated for each river. The forecasts for all of the gauging stations show significant skill (Table 6). Essentially, in both cases, the forecasts perform better than only using the median values. The forecast for the Thomson gauging station shows the greatest skill. Using the Thomson gauging station as an example, this result suggests that a 35 % improvement is expected over a decision based on the median flow of each month. Again using the Thomson as an example, Fig. 5 shows the forecast probability of flow exceeding median flow and the number of forecast successes and failures. The number of forecast successes and failures was calculated by comparing the outcome of Eq. (6) with whether the observed flow exceeded median flow or not.

$$\text{forecast flow}_i = \begin{cases} 1 & \text{if } p(\text{flow}_i > c) > 0.5 \\ 0 & \text{if } p(\text{flow}_i > c) \leq 0.5 \end{cases} \quad (6)$$

Also shown in Fig. 5. in gray are the cross-validation results for each of the 36 models. Importantly this shows that the cross-validation results are similar for each model.

A further important observation is that as the flow threshold increases, the value for BSS decreases (not shown) suggesting that as the flow threshold increases the system becomes less easy to forecast. A logical reason for this

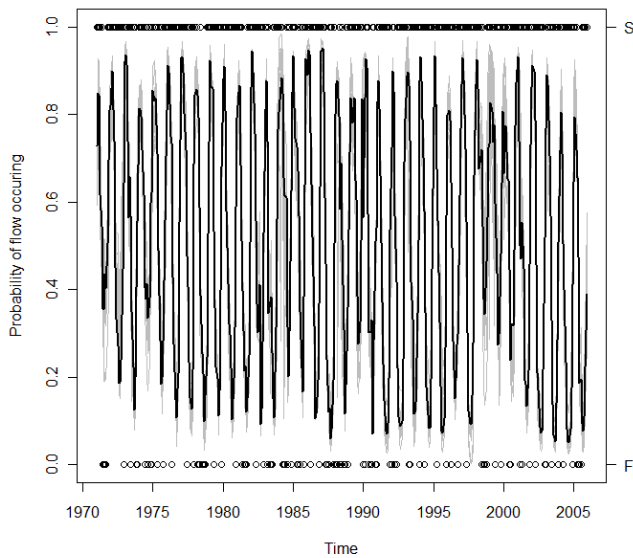


Fig. 5. Shows the probability of exceeding median flow for the Thomson River at Longreach. The circles indicate where the forecast has either succeeded (*S*) or failed (*F*). In this case 74 % of the forecasts were successful. The grey lines are the outputs from each of the cross-validated models.

observation is that at the higher flow thresholds the number of observed flows decreases, adding to the decrease in the forecast skill. In general, it appears that it is not possible to forecast the larger (extreme) flow events 12 months ahead. Rather, it is possible to predict wetter or dryer than average periods. Forecast skill tends to be higher for the 12 month ahead forecast as compared to the 6 month ahead forecast. This suggests the importance of the seasonal term in the forecast. Given that forecasts of streamflow are generally better than rainfall, our findings support findings of Westra and Sharma (2010) who show that global SST explain to explain a small percentage of rainfall variability at lags of 12 months.

From Table 6 and Eq. (5) it is possible to derive a forecast monthly flow duration curve 12 months ahead in time by generating regularly spaced flow threshold values up to a maximum threshold, say the maximum recorded flow (Fig. 6). The advantage of presenting forecasts as a flow duration curve is that they are already used by water managers to determine water extraction rates, irrigators for irrigation planning and by biologists to determine environmental flows (Acreman, 2005; Cigizoglu and Bayazit, 2000). Aside from the Thomson River, the forecast probability of flow is systematically overestimated for the other river systems. One reason for this is that the Box-Cox *t* distribution is not capturing all the skewness in these datasets and thus cannot generate the full range of probabilities. One potential solution is to use mixture distributions for the streamflow intensity (Stasinopoulos and Rigby, 2007) but this is not explored further.

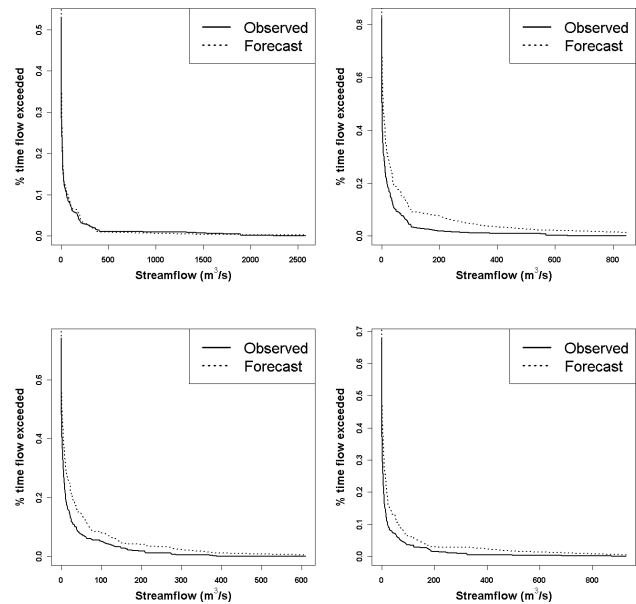


Fig. 6. Average monthly forecast and observed flow duration curve, Thomson River (top left panel), Bulloo River (top right panel), Paroo River (bottom left panel), and the Warrego River (bottom right panel).

4 General discussion

This study has demonstrated the ability of flexible statistical models to make skilful forecasts of intermittent streamflow in large catchments in inland Australia. In the absence of detailed understanding of complex large semi-arid catchments, statistical approaches, such as the demonstrated GAMLSS framework offer advantages over deterministic and conceptual catchment models for forecasts. From an explanatory view, the work has highlighted the influence of the Pacific Ocean SST of monthly flows in these catchments, increasing our understanding of these climatic drivers on Eastern Australian streamflow. It is also clear, however, that a single flexible sinusoidal term representing seasonality representing unknown periodicities explains much of the variability in these river systems. Additionally, the temporal covariate Time gives important explanations of long term trends such as the decrease in the observed Balonne flows due to water extraction for irrigation.

As this study is primarily a demonstration of a method, there is great scope for future work building on this approach for forecasting both streamflow and rainfall. For example, we have not considered antecedent soil moisture as a covariate in the model (Timbal et al., 2002) as this is relatively unworkable for the long range forecasts considered here. However, for shorter range forecasts, this could easily be introduced in the catchment process layer by incorporating a covariate based on the number of days or months from the start of a dry spell derived from local daily flow or rainfall records

Table 5. Intensity models for river systems in south western Queensland. The models for $\hat{\mu}$ (median), $\hat{\sigma}$ (scale parameter; approximately the coefficient of variation) are given. $\hat{\nu}$ (skewness) and $\hat{\tau}$ (kurtosis) are constants and are not given.

River	Lag (months)	Intensity Model
Thomson	6	$\hat{\mu} = 14.54 - s(\text{sine}) - s(\text{Niño}4) + s(\text{IOD})$ $\hat{\sigma} = 2.00 - s(\text{Niño}1.2)$
	12	$\hat{\mu} = 10.89 + s(\text{sine}) + 0.40 \text{ Niño}1.2 - s(\text{Niño}3.4)$ $\hat{\sigma} = -0.94 - 0.14 \text{ Niño}1.2 + 0.52 \text{ Niño}3 - 0.70 \text{ Niño}3.4$
Bulloo	6	$\hat{\mu} = 0.14 - s(\text{sine}) - s(\text{Niño}1.2) + 0.91 \text{ Niño}3 - s(\text{Niño}4)$ $\hat{\sigma} = 3.29 + s(\text{Niño}3) - 0.09 \text{ Niño}3.4$
	12	$\hat{\mu} = -4.36 + s(\text{sine}) + 0.59 \text{ Niño}1.2 - s(\text{Niño}3) + s(\text{Niño}3.4)$ $\hat{\sigma} = -0.22 + s(\text{Niño}3) - 0.08 \text{ IOD}$
Paroo	6	$\hat{\mu} = -1.22 - s(\text{sine}) - s(\text{Niño}1.2) + s(\text{Niño}3) - s(\text{Niño}4)$ $\hat{\sigma} = 2.51 + s(\text{Niño}3)$
	12	$\hat{\mu} = -1.48 + s(\text{sine}) + s(\text{Niño}1.2) - 1.59 \text{ Niño}3 + s(\text{Niño}3.4)$ $\hat{\sigma} = 0.96 + s(\text{sine})$
Warrego	6	$\hat{\mu} = -3.15 - s(\text{sine}) - s(\text{Niño}1.2) + 1.97 \text{ Niño}3 - 1.24 \text{ Niño}3.4$ $\hat{\sigma} = 2.83 + s(\text{Niño}3)$
	12	$\hat{\mu} = -1.95 + s(\text{sine}) + 0.70 \text{ Niño}1.2 - 2.09 \text{ Niño}3 + s(\text{Niño}3.4)$ $\hat{\sigma} = 0.96$
Balonne	6	$\hat{\mu} = -17.62 - s(\text{sine}) - s(\text{Time}) + 0.77 \text{ Niño}3 - 0.48 \text{ IOD}$ $\hat{\sigma} = -1.74 + s(\text{Time}) - s(\text{Niño}1.2)$
	12	$\hat{\mu} = -4.22 - s(\text{sine}) - s(\text{Time}) + 1.18 \text{ Niño}1.2 - 1.74 \text{ Niño}3 + s(\text{Niño}3.4) + s(\text{Niño}4)$ $\hat{\sigma} = 0.60 + 0.001 \text{ Time}$
Balonne Naturalised	6	$\hat{\mu} = 9.59 + s(\text{sine}) + 1.05 \text{ Niño}3 - 1.24 \text{ Niño}4 - s(\text{IOD})$ $\hat{\sigma} = 2.38 - s(\text{Niño}3)$
	12	$\hat{\mu} = -5.17 - s(\text{sine}) + 1.14 \text{ Niño}1.2 - s(\text{Niño}3) + s(\text{Niño}3.4)$ $\hat{\sigma} = -3.34 - 0.03 \text{ Niño}1.2 + 0.17 \text{ Niño}4$

Table 6. Skill of forecasting median flow at the five gauging stations and naturalised data.

Lag (months)	Score	Thomson	Bulloo	Paroo	Warrego	Balonne	Balonne naturalised
6	BSS	0.27	0.14	0.08	0.14	0.22	0.18
	ROC	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
	p-value						
12	BSS	0.36	0.14	0.12	0.17	0.22	0.18
	ROC	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
	p-value						

(Sharma and Lall, 1998). Furthermore, we have only used a small selection of available climate indices and we have only considered single lags of 6 and 12 months which could be extended to incorporate multiple lags or shorter lags for shorter range or seasonal forecasts. Examples of other indices which have been shown to be useful for forecasting precipitation or streamflow in Australia are the Tropical Indo-Pacific thermocline (Ruiz et al., 2007) and the Southern Annular Mode (Meneghini et al., 2007). The methodology can also be used

to identify temporal and spatial patterns in teleconnections between SST and precipitation or streamflow (Piechota et al., 1998; Wang et al., 2009). Similarly, the analysis could be extended to incorporate data from the entire global SST dataset (Sharma, 2000; Westra and Sharma, 2010). However, this will require a combination of balancing the degrees of freedom of the fitted splines and the number of covariates fitted, the use of a pre-selection methods for the covariates and an increase in the length of the datasets used. As an

extension, the proposed methodology can be used congruently with global climate models to translate forecast SST such as produced by POAMA (Alves et al., 2002) to local precipitation or streamflow. This application would also allow the inclusion of derived covariates which account for warming and thus potentially model the effect of warming on future streamflow data. Finally, only the binomial and Box-Cox t distributions have been considered in this study and it is expected that forecast will improve if other distributions are considered. In particular, it would be expected that using mixture distributions (Stasinopoulos and Rigby, 2007) for the intensity of streamflow will improve forecast skill. This is part of our ongoing research.

5 Conclusions

Using a GAMLSS regression framework it is possible to make a skilful forecast of the probability of monthly streamflow occurring 6 and 12 months ahead in highly variable intermittent streams in the inland regions of eastern Australia where only streamflow data is available. The GAMLSS framework is able to cope with non-linearity in the relationships between SST and monthly streamflow, which leads to superior model performance compared with more traditional linear models. Furthermore, in the absence of more detailed data and using synthetic covariates, it is possible to account for non-stationarity and seasonality in the data in an explanatory framework. The model output is probabilistic and hence the results can be presented as a probability of exceedance. This output can be used by irrigators, graziers and natural resource management staff to aid in decision making in these highly variable environments.

Acknowledgements. This research was conducted with the financial support from the Cotton Catchment Communities Cooperative Research Centre.

Edited by: J. Freer

References

- Abawi, Y., Dutta, S., Zhang, X., and McClymont, D.: ENSO-based streamflow forecasting and its application to water allocation and cropping decisions – an Australian experience, *Regional hydrological impacts of climate change – Impact assessment and decision making, Brazil*, 346–354, 2005.
- Acreman, M.: Linking science and decision-making: features and experience from environmental river flow setting, *Environ. Modell. Softw.*, 20, 99–109, 2005.
- Akaike, H.: A new look at the statistical model identification, *IEEE T. Automat. Contr.*, 19, 716–723, 1974.
- Alves, O., Wang, G., Zhong, A., Smith, N., Tzeitkin, F., Warren, G., Schiller, A., Godfrey, S., and Meyers, G.: POAMA: Bureau of Meteorology operational coupled model seasonal forecast system, *Proceedings of the ECMWF Workshop on the Role of the Upper Ocean in Medium and Extended Range Forecasting*, Reading, UK, 22–32, 2002.
- Barros, A. P. and Bowden, G. J.: Toward long-lead operational forecasts of drought: An experimental study in the Murray-Darling River Basin, *J. Hydrol.*, 357, 349–367, 2008.
- Barry, S. C. and Welsh, A. H.: Generalized additive modelling and zero inflated count data, *Ecol. Model.*, 157, 179–188, 2002.
- Berg, D.: Bankruptcy prediction by generalized additive models, *Appl. Stoch. Model. Bus.*, 23, 129–143, 2007.
- Braganza, K.: Seasonal climate summary southern hemisphere (autumn 2007): La Niña emerges as a distinct possibility in 2007, *Aust. Meteorol. Mag.*, 57, 191–198, 2008.
- Brown, B. G., Katz, R. W., and Murphy, A. H.: On the economic value of seasonal-precipitation forecasts: the following/planting problem, *B. Am. Meteorol. Soc.*, 67, 833–841, 1986.
- Buizza, R.: The value of probabilistic prediction, *Atmos. Sci. Lett.*, 9, 36–42, 2008.
- Cai, W. and Cowan, T.: Evidence of impacts from rising temperature on inflows to the Murray-Darling Basin, *Geophys. Res. Lett.*, 35, L07701, doi:10.1029/2008GL033390, 2008.
- Cai, W., Cowan, T., and Sullivan, A.: Recent unprecedented skewness towards positive Indian Ocean Dipole occurrences and its impact on Australian rainfall, *Geophys. Res. Lett.*, 36, L011705, doi:10.1029/2009GL037604, 2009.
- Chiew, F. H. S., Piechota, T. C., Dracup, J. A., and McMahon, T. A.: El Nino/Southern Oscillation and Australian rainfall, streamflow and drought: Links and potential for forecasting, *J. Hydrol.*, 192, 138–149, 1998.
- Chiew, F. H. S., Zhou, S. L., and McMahon, T. A.: Use of seasonal streamflow forecasts in water resources management, *J. Hydrol.*, 270, 135–144, 2003.
- Chiew, F. H. S., Teng, J., Vaze, J., Post, D. A., Perraud, J. M., Kirono, D. G. C., and Viney, N. R.: Estimating climate change impact on runoff across southeast Australia: Method, results, and implications of the modeling method, *Water Resour. Res.*, 45, W10414, doi:10.1029/2008WR007338, 2009.
- Chowdhury, S. and Sharma, A.: Multisite seasonal forecast of arid river flows using a dynamic model combination approach, *Water Resour. Res.*, 45, W10428, doi:10.1029/2008WR007510, 2009.
- Cigizoglu, H. K. and Bayazit, M.: A generalized seasonal model for flow duration curve, *Hydrol. Process.*, 14, 1053–1067, 2000.
- Collins, D.: Seasonal climate summary southern hemisphere (winter 1998): transition toward a cool episode (La Niña), *Aust. Meteorol. Mag.*, 48, 52–62, 1999.
- Cordery, I.: Long range forecasting of low rainfall, *Int. J. Climatol.*, 19, 463–470, 1999.
- Courtney, J.: Seasonal climate summary southern hemisphere (autumn 1998): decline of a warm episode (El Niño), *Aust. Meteorol. Mag.*, 47, 339–346, 1998.
- De Jong, P. and Heller, G. Z.: Generalized linear models for insurance data, in: *International Series on Actuarial Science*, edited by: Davis, M., Hylands, J., McCutcheon, J., Norberg, R., Panjer, H., and Wilson, A., Cambridge University Press, 2008.
- Dettinger, M. D. and Diaz, H. F.: Global characteristics of streamflow seasonality and variability, *J. Hydrometeorol.*, 1, 289–310, 2000.
- Dutta, S. C., Ritchie, J. W., Freebairn, D. M., and Abawi, Y.: Rainfall and streamflow response to El Nino Southern Oscillation: a case study in a semiarid catchment, Australia, *Hydrolog. Sci.*,

- 51, 1006–1020, 2006.
- Eilers, P. H. C. and Marx, B. D.: Flexible smoothing with B-splines and penalties, *Stat. Sci.*, 11, 89–101, 1996.
- Evans, J. L. and Allan, R. J.: El Niño/Southern Oscillation modification to the structure of the monsoon and tropical cyclone activity in the Australasian region, *Int. J. Climatol.*, 12, 611–623, 1992.
- Faraway, J. and Chatfield, C.: Time series forecasting with neural networks: a comparative study using the airline data, *Appl. Stat.*, 47, 231–250, 1998.
- Grunwald, G. K. and Jones, R. H.: Markov models for time series with mixed distribution, *Environmetrics*, 11, 327–339, 2000.
- Guisan, A., Edwards, T. C., and Hastie, T.: Generalized linear and generalized additive models in studies of species distributions: setting the scene, *Ecol. Model.*, 157, 89–100, 2002.
- Hameed, T. and Podger, G.: Use of the IQQM simulation model for planning and management of a regulated river system, *Integrated Water Resources Management*, edited by: Marino, M. A. and Simonovic, S. P., IAHS, 83–90, 2001.
- Hamill, T. M. and Wilks, D. S.: A probabilistic forecast contest and the difficulty in assessing short-range forecast uncertainty, *Weather Forecast.*, 10, 620–631, 1995.
- Hashino, T., Bradley, A. A., and Schwartz, S. S.: Evaluation of bias-correction methods for ensemble streamflow volume forecasts, *Evaluation*, 3, 561–594, 2006.
- Hastie, T. and Tibshirani, R.: Generalized additive models (with discussion), *Stat. Sci.*, 1, 297–310, 1986.
- Hayman, P., Crean, J., Mullen, J., and Parton, K.: How do probabilistic seasonal climate forecasts compare with other innovations that Australian farmers are encouraged to adopt?, *Aust. J. Agr. Res.*, 58, 975–984, 2007.
- Heller, G. Z., Stasinopoulos, D. M., Rigby, R. A., and Van Ogtrop, F. F.: Randomly stopped sum models: a hydrological application, 24th International Workshop on Statistical Modelling, New York, 2009.
- Hyndman, R. J. and Grunwald, G. K.: Generalized additive modelling of mixed distribution markov models with application to Melbourne’s rainfall, *Aust. NZ. J. Stat.*, 42, 145–158, 2000.
- Jolliffe, I. T. and Stephenson, D. B.: Forecast verification: a practitioner’s guide in atmospheric science, edited by: Jolliffe, I. T. and Stephenson, D. B., Wiley, 2003.
- Kiem, A. S. and Franks, S. W.: On the identification of ENSO-induced rainfall and runoff variability: a comparison of methods and indices, *Hydrolog. Sci. J./Journal des Sciences Hydrologiques*, 46, 715–728, 2001.
- Kiem, A. S. and Verdon-Kidd, D. C.: Climatic Drivers of Victorian Streamflow: Is ENSO the Dominant Influence?, *Aust. J. Water Resour.*, 13, 17–29, 2009.
- Krzysztofowicz, R.: Why should a forecaster and a decision maker use Bayes theorem, *Water Resour. Res.*, 19, 327–336, 1983.
- Krzysztofowicz, R.: The case for probabilistic forecasting in hydrology, *J. Hydrol.*, 249, 2–9, 2001.
- Kwiatkowski, D., Phillips, P., Schmidt, P., and Shin, Y.: Testing the null hypothesis of stationarity against the alternative of a unit root, *J. Econometr.*, 54, 78, 1992.
- Lachenbruch, P. A.: Comparisons of two-part models with competitors, *Stat. Med.*, 20, 1215–1234, 2001.
- Mason, S. and Graham, N.: Areas beneath the relative operating characteristics (roc) and relative operating levels (rol) curves: Statistical significance and interpretation, *Q. J. Roy. Meteorol. Soc.*, 128, 2145–2166, 2002.
- McAlpine, C. A., Syktus, J., Deo, R. C., Lawrence, P. J., McGowan, H. A., Watterson, I. G., and Phinn, S. R.: Modeling the impact of historical land cover change on Australia’s regional climate, *Geophys. Res. Lett.*, 34, L22711, doi:10.1029/2007GL031524, 2007.
- Meneghini, B., Simmonds, I., and Smith, I. N.: Association between Australian rainfall and the Southern Annular Mode, *Int. J. Climatol.*, 27, 109–121, 2007.
- Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P., and Stouffer, R. J.: Stationarity is dead: whither water management?, *Science*, 319, 573–574, 2008.
- Mjelde, J. W., Sonka, S. T., Dixon, B. L., and Lamb, P. J.: Valuing forecast characteristics in a dynamic agricultural production system, *Am. J. Agr. Econom.*, 70, 674–684, 1988.
- NCAR: Research Application Program: verification: Forecast verification utilities, R package version 1.31, 2010.
- Pappenberger, F. and Beven, K. J.: Ignorance is bliss: Or seven reasons not to use uncertainty analysis, *Water Resour. Res.*, 42, W05302, doi:10.1029/2005WR004820, 2006.
- Piechota, T. C. and Dracup, J. A.: Long-range streamflow forecasting using El Niño-Southern Oscillation indicators, *J. Hydrol. Eng.-ASCE*, 4, 144–151, 1999.
- Piechota, T. C., Chiew, F. H. S., Dracup, J. A., and McMahon, T. A.: Seasonal streamflow forecasting in eastern Australia and the El Niño-Southern Oscillation, *Water Resour. Res.*, 34, 3035–3044, 1998.
- Piechota, T. C., Chiew, F. H. S., Dracup, J. A., and McMahon, T. A.: Development of exceedance probability streamflow forecast, *J. Hydrol. Eng.-ASCE*, 6, 20–28, 2001.
- Pitman, A. J. and Perkins, S. E.: Regional projections of future seasonal and annual changes in rainfall and temperature over Australia based on skill-selected AR4 models, *Earth Interact.*, 12, 1–50, 2008.
- R Development Core Team: R: A language and environment for statistical computing, R Foundation for Statistical Computing, 2011.
- Rigby, R. and Stasinopoulos, D.: Generalized additive models for location, scale and shape, *J. Roy. Stat. Soc. C* 54, 507–554, 2005.
- Rigby, R. A. and Stasinopoulos, D. M.: Using the Box-Cox t distribution in GAMLSS to model skewness and kurtosis, *Stat. Modell.*, 6, 209–229, 2006.
- Roulin, E. and Vannitsem, S.: Skill of medium-range hydrological ensemble predictions, *J. Hydrometeorol.*, 6, 729–744, 2005.
- Ruiz, J. E., Cordery, I., and Sharma, A.: Forecasting streamflows in Australia using the tropical Indo-Pacific thermocline as predictor, *J. Hydrol.*, 341, 156–164, 2007.
- Saji, N. H. and Yamagata, T.: Structure of SST and surface wind variability during Indian Ocean dipole mode events: COADS observations, *J. Climate*, 16, 2735–2751, 2003.
- Schwarzer, G., Vach, W., and Schumacher, M.: On the misuses of artificial neural networks for prognostic and diagnostic classification in oncology, *Stat. Med.*, 19, 541–561, 2000.

- Sharma, A.: Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: Part 3 – A nonparametric probabilistic forecast model, *J. Hydrol.*, 239, 249–258, 2000.
- Sharma, A. and Lall, U.: A nonparametric approach for daily rainfall simulation, *Math. Comput. Simul.*, 48, 361–372, 1998.
- Sharma, A., Luk, K. C., Cordery, I., and Lall, U.: Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: Part 2 – Predictor identification of quarterly rainfall using ocean-atmosphere information, *J. Hydrol.*, 239, 240–248, 2000.
- Silverman, D. and Dracup, J. A.: Artificial neural networks and long-range precipitation prediction in California, *J. Appl. Meteorol.*, 39, 57–66, 2000.
- Simons, M., Podger, G., and Cooke, R.: IQQM—A hydrologic modelling tool for water resource and salinity management, *Environ. Softw.*, 11, 185–192, 1996.
- Soukup, T. L., Aziz, O. A., Tootle, G. A., Piechota, T. C., and Wulff, S. S.: Long lead-time streamflow forecasting of the North Platte River incorporating oceanic-atmospheric climate variability, *J. Hydrol.*, 368, 131–142, 2009.
- Stasinopoulos, D. and Rigby, R.: Generalized additive models for location scale and shape (GAMLSS) in R, *J. Stat. Softw.*, 23, 1–46, 2007.
- Thoms, M. C.: Floodplain-river ecosystems: lateral connections and the implications of human interference, *Geomorphology*, 56, 335–349, 2003.
- Thoms, M. C. and Parsons, M.: Identifying spatial and temporal patterns in the hydrological character of the Condamine-Balonne river, Australia, using multivariate statistics, *River Res. Appl.*, 19, 443–457, 2003.
- Timbal, B., Power, S., Colman, R., Viviand, J., and Lirola, S.: Does soil moisture influence climate variability and predictability over Australia?, *J. Climate*, 15, 1230–1238, 2002.
- Tootle, G. A. and Piechota, T. C.: Suwannee River long range streamflow forecasts based on seasonal climate predictors, *J. Am. Water Resour. Assoc.*, 40, 523–532, 2004.
- Trenberth, K. E. and Stepaniak, D. P.: Indices of El Niño evolution, *J. Climate*, 14, 1697–1701, 2001.
- Ummenhofer, C. C., England, M. H., McIntosh, P. C., Meyers, G. A., Pook, M. J., Risbey, J. S., Gupta, A. S., and Taschetto, A. S.: What causes southeast Australia's worst droughts?, *Geophys. Res. Lett.*, 36, L04706, doi:10.1029/2008GL036801, 2009.
- Underwood, F. M.: Describing long-term trends in precipitation using generalized additive models, *J. Hydrol.*, 364, 285–297, 2009.
- Verdon, D. C. and Franks, S. W.: Indian Ocean sea surface temperature variability and winter rainfall: Eastern Australia, *Water Resour. Res.*, 41, 335–345, 2005a.
- Verdon, D. C. and Franks, S. W.: Influence of Indian Ocean sea-surface temperature variability on winter rainfall across eastern Australia, *Regional Hydrological Impacts of Climatic Change – Impact assessment and decision making, Brazil*, 335–345, 2005b.
- Wang, C., Weisberg, R. H., and Virmani, J. I.: Western pacific interannual variability associated with the El Niño-Southern Oscillation, *J. Geophys. Res.*, 104, 5131–5149, 1999.
- Wang, G. and Hendon, H. H.: Sensitivity of Australian rainfall to Inter-El Niño variations, *J. Climate*, 20, 4211–4226, 2007.
- Wang, Q. J., Robertson, D. E., and Chiew, F. H. S.: A Bayesian joint probability modeling approach for seasonal forecasting of streamflows at multiple sites, *Water Resour. Res.*, 45, W05407, doi:10.1029/2008WR007355, 2009.
- Wang, S. Y., Gillies, R. R., Hipps, L. E., and Jin, J.: A transition-phase teleconnection of the Pacific quasi-decadal oscillation, *Clim. Dynam.*, 36, 1–13, 2010.
- Westra, S. and Sharma, A.: An Upper Limit to Seasonal Rainfall Predictability?, *J. Climate*, 23, 3332–3351, 2010.
- White, B.: The Australian experience, in: *Applications of seasonal climate forecasting in agriculture and natural ecosystems*, edited by: Hammer, G. L., Nicholls, N., and Mitchell, C., Kluwer, Dordrecht, 1–22, 2000.
- Whittingham, M. J., Stephens, P. A., Bradbury, R. B., and Freckleton, R. P.: Why do we still use stepwise modelling in ecology and behaviour?, *J. Animal Ecol.*, 75, 1182–1189, 2006.
- Wilks, D. S. and Murphy, A. H.: A decision-analytic study of the joint value of seasonal precipitation and temperature forecasts in a choice-of-crop problem, *Atmos.-Ocean*, 24, 353–368, 1986.
- Wilks, D. S.: *Statistical methods in the atmospheric sciences*, 3rd Edn., in: *International Geophysics*, edited by: Dmowska, R., Hartmann, D., and Rosby, T. H., Elsevier, Oxford, 2005.
- Wilks, D. S.: *Statistical methods in the atmospheric sciences*, in: *International Geophysics*, edited by: Dmowska, R., Hartman, D., and Rossby, T. H., Elsevier, San Diego, 2006.
- Wood, S. N.: *Generalized additive models An introduction with R*, in: *Texts in Statistical Science Series*, edited by: Carlin, B. P., Chatfield, C., Tanner, M., and Zidek, J., Chapman and Hall/CRC, Boca Raton, 2006.
- Wooldridge, S. A., Franks, S. W., and Kalma, J. D.: Hydrological implications of the Southern Oscillation: variability of the rainfall-runoff relationship, *Hydrolog. Sci. J.*, 46, 73–88, 2001.
- Yakowitz, S. J.: A stochastic model for daily river flows in an arid region, *Water Resour. Res.*, 9, 1271–1285, 1973.