Hydrology and Earth System Sciences Discussions

Improving statistical forecasts of seasonal streamflows using hydrological model output

D. E. Robertson, P. Pokhrel, and Q. J. Wang

CSIRO Land and Water, Highett, Victoria, Australia

Correspondence to: D. E. Robertson (david.robertson@csiro.au)

HESSD

9, 8701–8736, 2012

Improving statistical forecasts of seasonal streamflows

D. E. Robertson et al.

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

|◄ | ►|

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Abstract

Statistical methods traditionally applied for seasonal streamflow forecasting use predictors that represent the initial catchment condition and future climate influences on future streamflows. Observations of antecedent streamflows or rainfall commonly used to represent the initial catchment conditions are surrogates for the true source of predictability and can potentially have limitations. This study investigates a hybrid seasonal forecasting system that uses the simulations from a dynamic hydrological model as a predictor to represent the initial catchment condition in a statistical seasonal forecasting method. We compare the skill and reliability of forecasts made using the hybrid forecasting approach to those made using the existing operational practice of the Australian Bureau of Meteorology for 21 catchments in eastern Australia. We investigate the reasons for differences. In general, the hybrid forecasting system produces forecasts that are more skilful than the existing operational practice and as reliable. The greatest increases in forecast skill tend to be (1) when the catchment is wetting up but antecedent streamflows have not responded to antecedent rainfall, (2) when the catchment is drying and the dominant source of antecedent streamflow is in transition between surface runoff and base flow, and (3) when the initial catchment condition is near saturation intermittently throughout the historical record.

1 Introduction

Forecasts of streamflows for a range of forecast periods and lead times are valuable to a many users, including emergency services, hydroelectricity generators, irrigators, rural and urban water supply authorities and environmental managers. Forecasts of seasonal streamflows can inform tactical management of water resources, allowing water users and managers to plan operational water management decisions and assess the risks of alternative water use and management strategies. To be useful to

water users and managers in assessing risks, seasonal streamflow forecasts need to be accurate and reliably quantify forecast uncertainty.

Statistical methods are commonly used for operational seasonal streamflow forecasting around the world, due to their robustness and ability to reliably quantify forecast uncertainty (Plummer et al., 2009; Robertson and Wang, 2012; Garen, 1992; Pagano et al., 2009). Statistical streamflow forecasting methods use predictors that describe the two sources of seasonal streamflow predictability, the initial catchment condition and future climate influences (Robertson and Wang, 2012; Rosenberg et al., 2011). Climate indices, such as the Southern Oscillation Index or Indian Ocean Dipole Mode Index, are commonly used to represent the influence of future climate on streamflows (Robertson and Wang, 2012). The initial catchment condition is represented by observations of antecedent streamflow, antecedent rainfall or in cold climates snow depth or extent (Robertson and Wang, 2012; Garen, 1992). In all cases, the predictors used are simple indices that act as surrogates for the true source of predictability in a statistical model.

Antecedent streamflow or rainfall totals can be crude surrogate indicators of the initial catchment condition. Robertson and Wang (2012) found that a single predictor, selected from a pool of candidates that included antecedent streamflow and rainfall totals for up to the preceding three months, was sufficient to characterise the initial catchment conditions in the majority of locations and seasons. However, under some circumstances a second predictor could add additional independent information on the initial catchment condition. They concluded that more refined indicators of the initial catchment conditions could improve forecast skill.

Antecedent streamflow or rainfall totals are limited in their ability to provide a refined index describing initial catchment conditions for several reasons. Conceptually, catchment soil moisture and groundwater storages have upper and lower bounds. When these storages are full, streamflows (and rainfall) can continue to increase to levels beyond those that reflect catchment moisture storage. Therefore, when observed antecedent streamflow is very high, subsequent streamflow forecasts may be

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

considerably higher than the actual soil moisture or groundwater storage levels would cause. The dynamics of rainfall-runoff processes can also lead to antecedent streamflow or rainfall being a poor indicator of the initial catchment condition. When a catchment is wetting up, antecedent streamflows do not immediately respond to antecedent

5 rainfall, rather soil moisture and groundwater storages are replenished first. In this circumstance, antecedent streamflows can potentially underestimate the actual soil moisture conditions and lead to forecasts that are too low.

The use of dynamic hydrological models for seasonal streamflow forecasting has been investigated to overcome some of the limitations of statistical forecasting tech-

10 niques (for example, Bierkens and van Beek, 2009; Koster et al., 2010). Hydrological models describe the processes by which precipitation is converted into streamflow and in doing so explicitly represent catchment soil moisture and groundwater storages as state variables. Therefore, hydrological models can capture catchment dynamics that the simple indices used in statistical models cannot. When used in forecasting mode,

15 the condition of model state variables is initialised by running the model using observed forcing data up to the forecast date. A streamflow forecast is then produced by forcing the model with forecasts of rainfall and other forcing variables. Future rainfall is highly uncertain and difficult to accurately forecast, and therefore several sources of future rainfall have been investigated including conditional and unconditional historical cli-

20 mate sequences and output from seasonal climate forecasting models (Bierkens and van Beek, 2009; Wood et al., 2005). While these forecasts are derived from understanding of the hydrological processes occurring in the catchment, in many instances the direct forecasts from hydrological models are biased and do not reliably quantify forecast uncertainty (Shi et al., 2008).

25 Both statistical and dynamical streamflow forecasting methods appear to have strengths and weaknesses. Recently, Rosenberg et al. (2011) investigated the benefits of a hybrid seasonal forecasting system that uses the output from a physically based hydrological model as predictors in a statistical forecasting method in a climate where snow melt is the dominant source of streamflow. They showed that by using simulations

of snow water equivalent instead of observations as predictors that the skill of seasonal streamflow forecasts could be enhanced. The skill improvements were attributed to the simulations capturing the spatial and temporal variation in snow water equivalent better than the few sites that provide ground-based observations.

5   This paper also investigates a hybrid seasonal forecasting system, but in contrast to Rosenberg et al. (2011) we consider the problem in environments where snow melt is not an important source of streamflow. We investigate how the output of a dynamic hydrological model can be used to improve the representation of initial catchment conditions for statistical streamflow forecasting. We produce forecasts of three month 10 streamflow totals with the Bayesian joint probability (BJP) modelling approach (Wang and Robertson, 2011; Wang et al., 2009) using two alternative sets of predictors to represent initial catchment conditions. The first set of predictors represents the operational practice by the Bureau of Meteorology in Australia where the predictor with the highest Pseudo Bayes factor is selected from a pool of candidates comprising an- 15 tecedent streamflow and rainfall totals for up to the preceding three months (Robertson and Wang, 2012). The second set of predictors use simulations from a hydrological model that represents only the influence of initial catchment condition of streamflows for the forecast period. We compare the skill and reliability of these forecasts for 21 catchments in Eastern Australia and discuss the mechanisms by which the forecast 20 performance is improved.

2   Methods

2.1   Hydrological modelling

For this study, we use a hydrological model to produce simulations that represent only the influence of initial catchment conditions on seasonal streamflows. Hydrological 25 modelling is undertaken using WAPABA, a monthly water partition and balance model with two conceptual storages and five model parameters. WAPABA uses consumption

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

|◄ | ►|

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

curves to partition water according to supply and demand, which allow for spatial and temporal heterogeneity of catchment process. WAPABA has been shown to outperform other monthly models in Australia and simulate monthly streamflow volumes as well as daily models forced with daily data (Wang et al., 2011).

The WAPABA model parameters are calibrated by maximising a multi-objective function modified from Zhang et al. (2008). The model fit is evaluated using a uniformly weighted average of the Nash-Sutcliffe efficiency coefficient (Nash and Sutcliffe, 1970), the Nash-Sucliffe efficiency of the log transformed flows, the Pearson correlation coefficient and a symmetric measure of bias. Model calibration is performed using the Shuffled Complex Evolution algorithm (Duan et al., 1994).

Using calibrated model parameters, simulations are produced that represent only the initial catchment conditions influence on streamflow totals of the next three months. For a given date of interest, these simulations are obtained by running the model from the start of the historical record to the date of interest using observed forcing data, to initialise the model state variables, and then simulating streamflows for the subsequent three months using monthly climatology mean forcing data. A time series of these simulations of three month streamflow totals was produced by repeating the process for all months in the historical record. Using this approach, variation in the simulated three month streamflow totals for a given month is solely due to differences in the initial conditions of the soil moisture and groundwater storages and not related to variation in the climate forcing. Alternatives to using the monthly climatology mean forcing data were investigated, such as the climatology median forcing data and the mean and median of streamflow ensembles produced using all historically observed forcing data, but lead to final results that are no different to using the climatology mean forcing data.

2.2 Statistical streamflow forecasting

We use the Bayesian joint probability (BJP) modelling approach (Wang and Robertson, 2011; Wang et al., 2009) to produce joint forecasts of three month streamflow and rainfall totals. The BJP modelling approach assumes the joint distribution of forecast

variables and their predictors is described by a transformed multivariate normal distribution. A Yeo-Johnson transformation is for variables defined over the entire real space, while a log-sinh transformation (Wang et al., 2012) is used for variables that are defined for real values greater or equal to zero, for example streamflows or rainfall. Model parameters, including transformation parameters and reparameterisations of the means, variances and correlation coefficients of the multivariate normal distribution are inferred using Bayesian methods.

In this study, we primarily compare statistical streamflow forecasts made using two sets of predictors. The first set of predictors represents the existing operational practice of the Bureau of Meteorology in Australia. Predictors representing initial catchment conditions and future climate influences on streamflows are selected separately using the procedure described by Robertson and Wang (2012). The performance of a range of candidate predictors is assessed using the Pseudo Bayes factor (PsBF), a Bayes factor based on the cross-validation predictive density. The predictor with the highest PsBF is selected provided that the value is greater than a threshold that reduces the likelihood of choosing a predictor due to chance. Candidate predictors representing the initial catchment condition included antecedent streamflow and rainfall totals for up to the preceding three months and these are selected on their ability to forecast three month streamflow totals. Predictors representing climate during the forecast period are selected from a pool of 13 monthly climate indices lagged by up to three months and these are selected on their ability to forecast three month rainfall totals. At most two predictors are selected, one to represent the initial catchment condition and one to represent the climate during the forecast period. Forecasts of three month totals of streamflow and catchment average rainfall are made jointly. Separate models are established for each season and location to allow for inter-annual variations in climate and hydrological processes.

The second set or predictors used to make forecasts for this study replaces the selected predictors representing the initial catchment condition with the WAPABA simulations described in the previous section and total streamflow for the month preceding

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◀ | ▶

◀ | ▶

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

the forecast (lag-1 streamflow). The previous month's streamflow is included as a form of model updating to provide a real-time measure of the "true" condition of the catchment leading up to the forecast. The selected predictors representing climate during the forecast period are the same as in the first set of predictors.

2.3 Cross-validation for assessment of forecast performance

The hydrological and statistical modelling processes described in the preceding sections require observations to infer model parameters. For real-time forecasting applications, all available historical data of the appropriate quality may be used to infer model parameters. However if these model parameters are to be used to assess the performance of the forecasting methods for historical events, the forecast skill and reliability will be inflated. Therefore, it is necessary to assess forecast performance on data that has not been used for parameter inference and predictor selection. Traditionally, the skill of statistical forecasting models is assess using leave-one-out cross validation and this provides a realistic assessment of performance because the temporal sequence of data records is not preserved in model parameter inference. However, in this study we are also using a hydrological model which preserves and uses the temporal sequence of data records in model parameter inference, due to the presence of state variables in the model which carry information from one time step to the next. Therefore, forecast performance measures assessed using leave-one-out cross validation may be artificially inflated, because forecasts may not be independent of the data used for parameter inference. To limit this inflation of forecast performance measures we adopt a leave-one-plus-x-years-out cross validation approach. Ideally, the value of x is as small as possible to allow the data use to infer model parameters to reflect operational conditions in terms of available data length, while it needs to be sufficiently long to minimise any artificial inflation of forecast performance measures. For this study, we adopt leave-one-plus-four-years-out cross validation to assess forecast performance.

To make a cross-validation forecast for a year of interest, that year and the four subsequent years were removed from the data set used for model parameter inferences.

Hydrological model parameters were obtained by running the model for the entire record using all available forcing data, but omitting the observed streamflows for the year of interest and four subsequent years in the evaluation of the objective function. Simulations representing the initial catchment condition were produced for all years in the historical record and used in the statistical model to produce a forecast for the year of interest.

For selected predictors used in the statistical models, the predictors for the year of interest was selected using the PsBF computed using all events except the year of interest and the four subsequent years. Once model parameters were obtained, forecasts were made for the year of interest only and the process was repeated for all events in the historical record.

2.4 Forecast performance measures

There are many ways to assess the performance of streamflow forecasts. We assess the skill and reliability of the cross validation forecasts. Forecast skill is a measure of the quality of a set of forecasts relative to a baseline or reference set of forecasts (Jolliffe and Stephenson, 2003). We use skill scores that assess the percentage reduction in forecast error scores relative to the error scores of a reference forecast. For this study we assess forecast error using two scores; the root mean squared error in probability of the forecast median (RMSEP) (Wang and Robertson, 2011) and the continuous ranked probability score (CRPS). The reference forecasts used to compute the skill scores are the cross-validation distribution of historically observed (climatology) streamflows. The two skill scores adopted assess different aspects of the forecast distribution. The RMSEP skill score gives all forecast events similar opportunity to contribute to the overall assessment of forecast skill, but only considers the median of the forecast distribution. The CRPS skill score assesses the reduction in error of the whole forecast probability distribution, and can be sensitive to a few events with large forecast errors.

Forecast reliability measures assess the statistical consistency of the forecast probability distributions and the observed frequency of associated events (Toth et al., 2003).

For this study we use histograms of probability integral transforms (PIT) to assess the average reliability of the forecast probability distributions for all locations and seasons.

3   Catchments and data

In this study, we investigate the performance of forecasts made using the two differ-ent sets of predictors representing the initial catchment condition for 21 catchments in Eastern Australia that experience a range of climatic and hydrological conditions (Fig. 1 and Table 1). We use the observed monthly streamflow data obtained from various wa-ter resource management agencies and the Bureau of Meteorology. For most catch-ments, with the exception of some in Queensland and Victoria, the data are available from 1950 to 2008 (see Table 1). The monthly catchment average rainfall and potential evapotranspiration for each catchment are calculated from 5 km gridded data available from the Australian Water Availability Project (Jones et al., 2009). The monthly values of the 13 climate indices are obtained from the Bureau of Meteorology and described in Appendix 1.

4   Results

4.1   Forecast skill improvements

Forecasts made using WAPABA simulations and lag-1 streamflow as predictors to rep-resent initial catchment conditions generally have greater skill than forecasts made using selected predictors (Fig. 2). The average forecast skill score for forecasts made using the WAPABA simulations and lag-1 streamflow as predictors is 2.7 % greater for both the RMSEP and CRPS skill scores than for forecasts made using the selected pre-dictors. While this average improvement appears to be small the range of changes in both skill scores extends from −10 % to +25 %. The greatest increases and decreases

in forecast skill tend to occur for those locations and seasons where the skill of forecasts made using selected predictors is less than 10 %.

Figure 3 presents the increase in forecast skill that is achieved by replacing the selected predictors by the WAPABA simulations and lag-1 streamflow arranged by catchment and season. Increases in both skill scores are most pronounced in the Queensland catchments (at the top of Fig. 3) and for the MJJ, JJA, NDJ and DJF seasons in the Central Victorian and Upper Murray catchments. Decreases in forecast skill are most evident for the MAM and AMJ seasons in Central Victorian and Upper Murray catchments. The seasons where there is the greatest increase in skill tend to be those that cover the steepest rise or fall of the annual hydrograph and therefore this suggests that the selected predictors are unable to adequately capture the inter-annual variations in the dynamics of catchments wetting and drying.

Figure 4 presents the skill scores of cross validation forecasts by catchment and season. There is a distinct seasonal pattern in forecast skill, particularly for the catchments in the Upper Murray, Central and Southern Victoria. The highest skill forecasts tend to be for seasons that cover the period when the annual hydrograph is falling, while the lowest skill forecasts tend to be for seasons that cover the period when the annual hydrograph is rising. The Tasmanian catchments have low forecast skill year round because these catchments tend to remain near saturation all year, streamflows are strongly related to concurrent rainfall, and seasonal rainfall is difficult to forecast. The Queensland catchments are in tropical and sub-tropical environments with pronounced wet and dry seasons. Forecast skill tends to be highest for the dry seasons and lowest for seasons covering the wettest months (November to March). Skill also tends to be low for seasons and catchments where frequently rivers cease to flow. In these circumstances the forecast error, particularly the RMSEP error, of a climatology forecast is small and therefore it is difficult to further reduce forecast error.

4.2 Forecast reliability

Replacing the selected predictors representing initial catchment conditions with WA-PABA simulations and lag-1 streamflow produced little change in the reliability of streamflow forecasts. Figure 5 presents histograms of the PIT values for forecasts of made using both sets of predictors. The differences between the two histograms are small and the general pattern of the histograms is similar. Perfectly reliable forecasts will produce a PIT histogram that is a uniform distribution. Figure 5 suggests that when viewed collectively the forecasts are not necessarily reliable, with the most obvious deviations from uniformity occurring in the highest and lowest bins of the histogram. However when the reliability is assessed for each season and catchment separately, deviations from uniformity are within the range expected by sample variability.

4.3 Reasons for improvements in forecast skill

The improvements in forecast skill occur under three main sets of conditions: (1) when the catchment is wetting up but antecedent streamflows have not responded to antecedent rainfall; (2) when the catchment is drying and the dominant source of antecedent streamflow is in transition between surface runoff and base flow; and (3) when the initial catchment condition is near saturation intermittently throughout the historical record. Here we examine some examples of how replacing the selected predictors with WAPABA simulations and lag-1 streamflows influences cross-validation forecasts and improves forecast skill.

4.3.1 When the catchment is wetting up

The selected predictors are either antecedent streamflow or rainfall totals for the previous month. When the catchment is wetting up, antecedent streamflows are primarily base flow and have not necessarily responded to antecedent rainfall. Therefore, antecedent streamflow does not necessarily represent the wetness of the catchment well.

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◀ | ▶

◀ | ▶

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Where antecedent rainfall totals have been insufficient to saturate the catchment, they will primarily reflect the surface moisture conditions of the catchment.

Figure 6 provides an example of forecasts made for this situation for March-April-May forecasts for Kiewa River inflows to the Murray River. For this example, replacing the selected predictor representing initial catchment conditions with WAPABA simulations and lag-1 streamflow increases the RMSEP skill score from 9 % to 18 % and the CRPS skill score from 1 % to 10 %. The selected predictor representing initial catchment conditions is predominantly total streamflow for January and February, which will provided an indication of base flow conditions. Overall the forecast quantile ranges for a given forecast median are similar using both sets of predictors, however forecast medians are rearranged. By replacing the selected predictors with WAPABA simulations and lag-1 streamflow the forecast error is reduced, particularly for forecasts associated with observations in the upper and lower quartiles of the historical distribution (light grey shade in Fig. 6). The primary reason for the difference between the forecasts produced using the two sets of predictors is due the WAPABA simulations being more strongly correlated to streamflows during the forecast period than any of the candidate predictors representing initial catchment conditions used for predictor selection. The correlation between the WAPABA simulations and streamflows during the forecast period is also better preserved for independent cross validation events. This suggests that the WAPABA simulations provide a better representation of the process of catchment wetting up than any of the candidates.

4.3.2 When the catchment is drying out

When the catchment is drying out, antecedent streamflows may be dominated by direct surface runoff if there has been recent rain or by base flows if there has not been. Total monthly streamflows of similar magnitude can be produced by both sources and therefore antecedent streamflow may not necessarily provide the best indicator of the wetness of a catchment. Figure 7 provides an example of forecasts made for this situation

of November-December-January forecasts for inflows into Dartmouth Reservoir. For this example, replacing the selected predictor representing initial catchment conditions with WAPABA simulations and lag-1 streamflow increases the RMSEP skill score from 19 % to 31 % and the CRPS skill score from 17 % to 28 %. The selected predictor representing initial catchment conditions is predominantly total streamflow for September and October. Like the previous example when the catchment is wetting up, the forecast quantile ranges for a given forecast median are similar using both sets of predictors and the forecast medians are rearranged. However, in contrast to the previous example, the skill gains are achieved by reducing the errors of the forecast median for events corresponding to observations in the central quartiles (mid grey shade in Fig. 7) of the historical observations rather than in the outer quartiles. It is for these moderate events that antecedent streamflows could be sourced from either surface runoff or base flow and the WAPABA simulations can distinguish the dominant source, whereas the candidates used in predictor selection cannot. As with the previous example the WAPABA simulations are more strongly correlated to streamflows during the forecast period than any of the candidate predictors representing initial catchment conditions used in the predictor selection.

4.3.3 When the catchment is intermittently saturated

The soil moisture and groundwater stores of a catchment are bounded, that is soil can become saturated and groundwater water tables can approach the surface. However, the antecedent streamflow and rainfall totals used as candidate indicators of the catchment condition in predictor selection are theoretically unbounded, that is they continue to increase when the soil moisture and groundwater stores are full. When the soil in a catchment is saturated and groundwater stores are near capacity in the month preceding a forecast, antecedent streamflow and rainfall are poor indicators of the condition. For a given forecast period, a catchment may be saturated consistently or intermittently throughout the historical record. For much of the year, the Tasmanian

catchments considered in this study provide examples of consistently saturated catchment conditions throughout the historical record. For these locations and seasons the forecast skill is close to zero and replacing the selected predictors with WAPABA simulations and lag-1 streamflow results in little change in forecast skill.

5   When the catchment is intermittently saturated through the historical record, replacing the selected predictors with WAPABA simulations and lag-1 streamflows can improve the skill of streamflow forecasts. Figure 8 provides an example of when the catchment is intermittently saturated through the historical record using forecasts of July-August-September inflows into Upper Yarra Reservoir. For this example, replacing
10 the selected predictor representing initial catchment conditions with WAPABA simulations and lag-1 streamflow increases the RMSEP skill score from 5 % to 17 % and the CRPS skill score from 3 % to 12 %. The selected predictor representing initial catchment conditions is predominantly total streamflow for April, May and June. Like the previous examples, the forecast quantile ranges for a given forecast median are similar
15 using both sets of predictors and the forecast medians are rearranged to more closely match to the observations. The skill gains are achieved by reducing errors in the forecast median for all events throughout the entire range of historical observations. This is due to streamflows during the forecast period being more strongly correlated with the WAPABA simulations that any of the candidate indicators of initial catchment conditions
20 considered in the predictor selection process.

      The relationship between streamflow during the forecast period and the WAPABA simulations is approximately linear. The relationship between streamflows during the forecast period and other variables used as candidates for predictor selection appears linear for low values of the candidate predictor and deviate from linearity above
25 a threshold value. This two part relationship suggests that for low values the candidate predictors are reasonable indicators of the initial catchment conditions, but at higher values they are not. Examining the initialised state variables of WAPABA used in producing the simulations that represent initial catchment conditions suggests that when the antecedent streamflows exceed this threshold the soil moisture store is at or near

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

|◄ | ►|

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

CC BY

capacity and the groundwater store level is very high. Therefore, the improvements in forecast skill arising from replacing the selected predictors with WAPABA simulations and lag-1 streamflow can be attributed to a better representation of the catchment process when the catchment intermittently becomes very wet.

5 Discussion

In this study, selected predictors representing initial catchment conditions were replaced by a combination of simulations from a dynamic hydrological model and lag-1 streamflows. Lag-1 streamflows were included as a form of model updating to provide a real-time measure of the actual catchment condition leading up to the forecast. However, as lag-1 streamflow is not always a good indicator of the catchment condition and its inclusion may moderate some of the benefit of using the WAPABA simulations. Figure 9 presents the increases in forecast skill arising from using lag-1 streamflows as well as WAPABA simulations to represent the initial catchment condition. In general, including lag-1 streamflow to provide a real-time measure of the actual catchment condition has little impact on the forecast skill. In some instances including lag-1 streamflow increases forecast skill for a small number of seasons and locations, but most importantly it does not degrade forecast skill. Therefore, it appears that including lag-1 streamflows as a form of model update is appropriate.

For the majority of catchments and seasons the skill of forecasts is due to the knowledge of the initial catchment condition. Figure 10 illustrates the contribution of the climate indices to forecast skill. When the points are in Fig. 10 are located on the 1 : 1 line the climate indices make no contribution to the skill of streamflow forecasts, while points below the 1 : 1 line suggest that climate indices improve forecast skill. The contribution of climate indices to streamflow forecast skill tends to be largest for catchments in Queensland, where there is the strongest evidence for using climate indices to forecast seasonal rainfall (Schepen et al., 2012; Wang et al., 2011).

The points above the 1 : 1 line in Fig. 10 indicate that forecasts made without se-lected climate indices are more skilful that those made with climate indices. This sug-gests that while there is evidence for using climate indices to forecast rainfall during a fitting period, the fitted relationship does not perform well for independent forecasts.

5 The approach to assessing forecast skill used in this paper is designed to expose cir-cumstances where this occurs and assess the true skill of the predictor selection and forecasting approaches by using cross validated predictors as well as cross-validated model parameters. Where the predictors are not cross validated, it is likely that the reported forecast skill is artificially inflated and will not be maintained in operational

10 applications (Michaelsen, 1987; DelSole and Shukla, 2009).

For this study, we adopted leave-one-plus-four-years-out cross validation to assess forecast performance. We adopted this approach to limit the potential for forecast per-formance to be artificially inflated due to the state variables in WAPABA carrying infor-mation from one time step to the next and resulting in forecasts that are not indepen-

15 dent of data used for parameter inference. We tested the assumption that leave-one-plus-four-years-out was sufficient to create independent forecasts by also assessing forecast skill using leave-one-plus-one-years-out and leave-one-plus-nine-years-out. In the assessment, we fixed the climate predictors so that variations in the forecast perfor-mance measures were solely due to the different periods omitted from the data used

20 for parameter inference. The differences between the forecast skill scores produced using the different cross validation methods tended to be within the range of sample variability (not shown). Where there were differences there was no clear pattern to the best or worst performing cross validation approach and therefore the adopted approach appears appropriate.

25 The climate indices used to represent the future climate influences on streamflow in this study are surrogates for the true source of climate predictability. The true source of climate predictability arises from understanding of the initial conditions of the ocean, atmosphere and land surface and the processes by which these conditions evolve and interact. Many dynamic coupled ocean–atmosphere models have been developed

to produce seasonal climate forecasts (for example: Alves et al., 2002). These models simulate the dynamic evolution of chaotic ocean and atmospheric processes from estimates of the ocean, atmosphere and land surface initial conditions. Forecasts of rainfall, or other atmospheric variables, produced by these models may provide bet-
5 ter indicators of future climate influences on seasonal streamflows than simple climate indices because they integrate a wide range of initial conditions. They also provide the opportunity for the use of concurrent relationships, which tend to be stronger than lagged relationships. However, comprehensive analysis of dynamic climate model output is necessary to better understand the quality of the forecasts and which variables
10 are useful for streamflow forecasting. Future work will investigate the use of such output for seasonal forecasting of streamflows in Australia using statistical models and rainfall-runoff models.

WAPABA simulates monthly streamflow totals in validation periods using monthly forcing data as well as daily rainfall-runoffs models forced with daily data (Wang et al.,
15 2011). However, the skill of raw WAPABA simulations representing the initial catchment condition was considerably poorer than the forecasts resulting from using the WAPABA simulations as a predictor in the BJP modelling approach. The poor skill of the raw WAPABA simulations representing initial catchment condition is primarily due to variation in seasonal biases that overall forecast performance measure do not diagnose (not
20 shown). The BJP modelling approach was able to extract information from the biased WAPABA simulations and produce skilful forecasts with minimal biases.

The water balance model used in this study is a relatively simple, lumped monthly model. Situations may exist where such a model may not necessarily provide sufficient spatial, temporal or process resolution to adequately describe the catchment condition
25 at the forecast time. In these situations, more sophisticated models may be warranted to describe to describe the catchment conditions. Simulations from more sophisticated models can also be included as predictors in the BJP modelling approach using the process described in this paper.

Title Page

| Abstract | Introduction |

| Conclusions | References |

| Tables | Figures |

◄◄   ►►

◄   ►

Back   Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

6   Conclusions

Forecasts of seasonal streamflows are valuable to a wide range of users. Traditionally, these forecasts are produced using statistical methods with observations of antecedent streamflows or rainfall as predictors to represent the condition of the catchment at the forecast date and with climate indices as predictors to represent the influence of future climate. These predictors are surrogates for the true source of predictability and can potentially have limitations. Dynamic hydrological models have also been used for streamflow forecasting, but often require statistical post-processing to remove biases and correct the reliability of forecast probability distributions. This study has investigated whether a hybrid seasonal forecasting system that uses the output of a dynamic hydrological model as a predictor in a statistical forecasting approach can lead to more skilful forecasts. Forecasts of three month streamflow totals were made using two alternative sets of predictors to represent initial catchment conditions: predictors selected using the method employed in the operational practice by the Bureau of Meteorology in Australia; and the combination of simulations from a monthly water balance model that represents the influence of initial catchment condition of streamflows and lag-1 streamflow. The skill and reliability of streamflow forecasts made using these sets of predictors were compared for 21 catchments in Eastern Australia and insights into the reasons for any differences investigated.

In general, replacing selected predictors representing the initial catchment condition with simulations from a monthly balance model and lag-1 streamflow increases the forecast skill and has little impact on forecast reliability. The magnitude of the skill increases varies with location and season. The greatest increases in forecast skill tend to be for three sets of circumstances: (1) when the catchment is wetting up but antecedent streamflows have not responded to antecedent rainfall; (2) when the catchment is drying and the dominant source of antecedent streamflow is in transition between surface runoff and base flow; and (3) when the initial catchment condition is near saturation intermittently throughout the historical record. There is little change in forecast skill for

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

catchments and seasons that are very dry or consistently saturated throughout the historical record. Even with the skill improvements realised by replacing the selected predictors, the skill of streamflow forecasts tends to be the highest for seasons that include the falling limb of the annual hydrograph, when seasonal streamflows are strongly related to the initial catchment condition. The skill tends to be the lowest for seasons that include the rising limb, when seasonal streamflows are strongly related to concurrent rainfall. In general the contribution of climate indices used to represent the influence of future climate to forecast skill is small but comparable to that of forecasts of seasonal rainfall. Future work will investigate how using the output of dynamic climate models may improve this situation.

Lag-1 streamflow was included as a predictor in addition to the monthly water balance simulations as a form of model updating to provide a real-time measure of the catchment condition. In general, it contributes little to forecast skill, but for some seasons and location skill increases of up to 20 % are realised by its inclusion. Most importantly, including lag-1 streamflow does not degrade forecast skill and therefore can be confidently included as a predictor for operational forecasts. The use of a more sophisticated hydrological model with increased spatial, temporal or process resolution may reduce the need for model updating. The output of such a higher resolution hydrological model could be used as a predictor in the BJP modelling approach using the methods described in this paper.

Appendix A

Climate indices used as candidate predictors to represent the influence of climate during the forecast period on streamflows (see Table A1).

10 **References**

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, National Drought Forum, Brisbane, 2002.

Ashok, K., Guan, Z. Y., and Yamagata, T.: Influence of the Indian Ocean Dipole on the Australian
15 winter rainfall, Geophys. Res. Lett., 30, 1821, doi:10.1029/2003GL017926, 2003.

Bierkens, M. F. P. and van Beek, L. P. H.: Seasonal predictability of European discharge: NAO and hydrological response time, J. Hydrometeorol., 10, 953–968, doi:10.1175/2009jhm1034.1, 2009.

DelSole, T. and Shukla, J.: Artificial skill due to predictor screening, J. Climate, 22, 331–345,
20 doi:10.1175/2008jcli2414.1, 2009.

Duan, Q., Sorooshian, S., and Gupta, V. K.: Optimal use of the SCE-UA global optimization method for calibrating watershed models, J. Hydrol., 158, 265–284, 1994.

Garen, D. C.: Improved techniques in regression-based streamflow volume forecasting, J. Water Res. Pl.-ASCE, 118, 654–670, doi:10.1061/(ASCE)0733-9496(1992)118:6(654), 1992.

25 Jolliffe, I. T. and Stephenson, D. B.: Forecast Verification: a Practitioner's Guide in Atmospheric Science, J. Wiley, Chichester, 240 pp., 2003.

Jones, D. A., Wang, W., and Fawcett, R.: High-quality spatial climate data-sets for Australia, Aust. Meteorol. Oceanogr. J., 58, 233–248, 2009.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R., and Joseph, D.: The NCEP/NCAR 40-year reanalysis project, B. Am. Meteorol. Soc., 77, 437–471, 1996.

5 Koster, R. D., Mahanama, S. P. P., Livneh, B., Lettenmaier, D. P., and Reichle, R. H.: Skill in streamflow forecasts derived from large-scale estimates of soil moisture and snow, Nat. Geosci., 3, 613–616, 2010.

Marshall, G. J.: Trends in the southern annular mode from observations and reanalyses, J. Climate, 16, 4134–4143, doi:10.1175/1520-0442(2003)016<4134:TITSAM>2.0.CO;2, 2003.

10 Michaelsen, J.: Cross-validation in statistical climate forecast models, J. Clim. Appl. Meteorol., 26, 1589–1600, 1987.

Mo, K. C.: Relationships between low-frequency variability in the Southern Hemisphere and sea surface temperature anomalies, J. Climate, 13, 3599–3610, doi:10.1175/1520-0442(2000)013<3599:RBLFVI>2.0.CO;2, 2000.

15 Murphy, B. F. and Timbal, B.: A review of recent climate variability and climate change in Southeastern Australia, Int. J. Climatol., 28, 859–879, doi:10.1002/joc.1627, 2008.

Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I – a discussion of principles, J. Hydrol., 10, 282–290, doi:10.1016/0022-1694(70)90255-6, 1970.

Pagano, T. C., Garen, D. C., Perkins, T. R., and Pasteris, P. A.: Daily updating of operational
20 statistical seasonal water supply forecasts for the Western US, J. Am. Water Resour. As., 45, 767–778, doi:10.1111/j.1752-1688.2009.00321.x, 2009.

Plummer, N., Tuteja, N. K., Wang, Q. J., Wang, E., Robertson, D. E., Zhou, S., Schepen, A., Alves, O., Timbal, B., and Puri, K.: A seasonal water availability prediction service: Opportunities and Challenges, in: 18th World IMACS/MODSIM Congress, Cairns, 13–17 July 2009,
25 2009.

Risbey, J. S., Pook, M. J., McIntosh, P. C., Wheeler, M. C., and Hendon, H. H.: On the remote drivers of rainfall variability in Australia, Mon. Weather Rev., 137, 3233–3253, doi:10.1175/2009mwr2861.1, 2009.

Robertson, D. E. and Wang, Q. J.: A bayesian approach to predictor selection for seasonal
30 streamflow forecasting, J. Hydrometeorol., 13, 155–171, 2012.

Rosenberg, E. A., Wood, A. W., and Steinemann, A. C.: Statistical applications of physically based hydrologic models to seasonal streamflow forecasts, Water Resour. Res., 47, W00H14, doi:10.1029/2010wr010101, 2011.

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

|◄ | ►|

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Ruiz, J. E., Cordery, I., and Sharma, A.: Impact of Mid-Pacific Ocean thermocline on the prediction of Australian rainfall, J. Hydrol., 317, 104–122, doi:10.1016/j.jhydrol.2005.05.012, 2006.

Saji, N. H., Goswami, B. N., Vinayachandran, P. N., and Yamagata, T.: A dipole mode in the Tropical Indian Ocean, Nature, 401, 360–363, 1999.

Schepen, A., Wang, Q. J., and Robertson, D. E.: Evidence for using climate indices to forecast Australian seasonal rainfall, J. Climate, 25, 1230–1246, 2012.

Shi, X., Wood, A. W., and Lettenmaier, D. P.: How essential is hydrologic model calibration to seasonal streamflow forecasting?, J. Hydrometeorol., 9, 1350–1363, doi:10.1175/2008JHM1001.1, 2008.

Smith, T. M., Reynolds, R. W., Peterson, T. C., and Lawrimore, J.: Improvements to NOAA's historical merged land-ocean surface temperature analysis (1880–2006), J. Climate, 21, 2283–2296, doi:10.1175/2007jcli2100.1, 2008.

Toth, Z., Talagrand, O., Candille, G., and Zhu, Y.: Probability and ensemble forecasts, in: Forecast Verification: a Practitioner's Guide in Atmospheric Science, edited by: Jolliffe, I. T. and Stephenson, D. B., J. Wiley, Chichester, 137–164, 2003.

Troup, A. J.: Southern oscillation, Q. J. Roy. Meteorol. Soc., 91, 490–506, doi:10.1002/qj.49709139009, 1965.

Verdon, D. C. and Franks, S. W.: Indian Ocean sea surface temperature variability and winter rainfall: Eastern Australia, Water Resour. Res., 41, W09413, doi:10.1029/2004WR003845, 2005.

Wang, Q. J. and Robertson, D. E.: Multisite probabilistic forecasting of seasonal flows for streams with zero value occurrences, Water Resour. Res., 47, W02546, doi:10.1029/2010wr009333, 2011.

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, Q. J., Pagano, T. C., Zhou, S. L., Hapuarachchi, H. A. P., Zhang, L., and Robertson, D. E.: Monthly versus daily water balance models in simulating monthly runoff, J. Hydrol., 404, 166–175, 2011.

Wang, Q. J., Shrestha, D. L., Robertson, D. E., and Pokhrel, P.: A log-sinh transformation for data normalisation and variance stabilisation, Water Resour. Res., 48, W05514, doi:10.1029/2011WR010973, 2012.

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◄ | ►

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Wood, A. W., Kumar, A., and Lettenmaier, D. P.: A retrospective assessment of National Centers for Environmental Prediction climate model-based ensemble hydrologic forecasting in the Western United States, J. Geophys. Res.-Atmos., 110, D04105, doi:10.1029/2004jd004508, 2005.

5 Zhang, L., Potter, N., Hickel, K., Zhang, Y., and Shao, Q.: Water balance modeling over variable time scales based on the Budyko framework – model development and testing, J. Hydrol., 360, 117–131, 2008.

Improving statistical forecasts of seasonal streamflows

D. E. Robertson et al.

Title Page

Abstract	Introduction
Conclusions	References
Tables	Figures

|◄ | ►|

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Table 1. Attributes of the 21 catchments used for the study (Res. indicates Reservoir inflow, HES indicates inflow to hydroelectric scheme).

ID	Catchment	Region	Available Record	Catchment area (km^2)	Mean annual rainfall (mm)	Mean annual flow (mm)	Annual runoff coeff.
1	Barron River	Queensland	1950–2008	228	1367	605 (138 GL)	0.44
2	South Johnstone River	Queensland	1974–2008	390	3128	2018 (787 GL)	0.65
3	Burdekin River	Queensland	1967–2008	36260	567	76 (2765 GL)	0.13
4	Brisbane River	Queensland	1950–2000	3866	846	79 (304 GL)	0.09
5	Somerset Res.	Queensland	1950–2000	1366	1245	289 (395 GL)	0.23
6	Hume Res.	Upper Murray	1950–2008	12184	819	227 (2764 GL)	0.28
7	Dartmouth Res.	Upper Murray	1950–2008	3193	1042	279 (890 GL)	0.27
8	Kiewa River	Upper Murray	1965–2008	1748	1099	248 (433 GL)	0.23
9	Ovens River	Upper Murray	1959–2008	7543	963	175 (1320 GL)	0.18
10	Nillahcootie Res.	Central Victoria	1950–2008	422	942	150 (63 GL)	0.16
11	Eildon Res.	Central Victoria	1950–2008	3877	1104	373 (1447 GL)	0.34
12	Goulburn Res.	Central Victoria	1950–2008	7166	769	188 (1349 GL)	0.24
13	Eppalock Res.	Central Victoria	1950–2008	1749	630	98 (172 GL)	0.16
14	Cairn Curran Res.	Central Victoria	1950–2008	1603	617	72 (115 GL)	0.12
15	Tullaroop Res.	Central Victoria	1950–2008	702	633	77 (54 GL)	0.12
16	Thompson Res.	Southern Victoria	1950–2008	487	1299	485 (236 GL)	0.37
17	Upper Yarra Res.	Southern Victoria	1950–2008	336	1387	443 (149 GL)	0.32
18	Maroondah Res.	Southern Victoria	1950–2008	129	1351	577 (74 GL)	0.43
19	O'Shannassy Res.	Southern Victoria	1950–2008	127	1404	766 (97 GL)	0.55
20	Mersey-Forth HES	Tasmania	1950–2008	2698	1900	793 (2141 GL)	0.42
21	King HES	Tasmania	1950–2008	731	2703	1724 (1260 GL)	0.64

Table A1. Climate indices included as candidate predictors of seasonal forecast of streamflow and data sources.

Candidate predictors	Period of record used	Data source
Southern Oscillation Index (SOI) (Troup, 1965)	1950–2008	Australian Bureau of Meteorology
NINO3 (SST anomaly over 90° W–150° W, 5° S–5° N)	1950–2008	NCAR, ERSST.v3 (Smith et al., 2008)
NINO3.4 (SST anomaly over 120° W–170° W, 5° S–5° N)	1950–2008	NCAR, ERSST.v3 (Smith et al., 2008)
NINO4 (SST anomaly over 150° W–160° E, 5° S–5° N)	1950–2008	NCAR, ERSST.v3 (Smith et al., 2008)
ENSO Modoki Index (Ashok et al., 2003)	1950–2008	NCAR, ERSST.v3 (Smith et al., 2008)
20° Isotherm (Ruiz et al., 2006)	1980–2008	Bureau of Meteorology
Indian Ocean Dipole Mode Index (Saji et al., 1999)	1950–2008	NCAR, ERSST.v3 (Smith et al., 2008)
Indian Ocean West Pole Index (Saji et al., 1999)	1950–2008	NCAR, ERSST.v3 (Smith et al., 2008)
Indian Ocean East Pole Index (Saji et al., 1999)	1950–2008	NCAR, ERSST.v3 (Smith et al., 2008)
Indonesia Index (Verdon and Franks, 2005)	1950–2008	NCAR, ERSST.v3 (Smith et al., 2008)
Tasman Sea Index (Murphy and Timbal, 2008)	1950–2008	NCAR, ERSST.v3 (Smith et al., 2008)
Southern Annular Mode (Marshall, 2003)	1979–2008	Antarctic Oscillation Index NOAA (Mo, 2000)
140° E Blocking Index (Risbey et al., 2009)	1950–2008	Calculated from NCEP/NCAR reanalysis data (Kalnay et al., 1996)

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◀◀ | ▶▶

◀ | ▶

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion
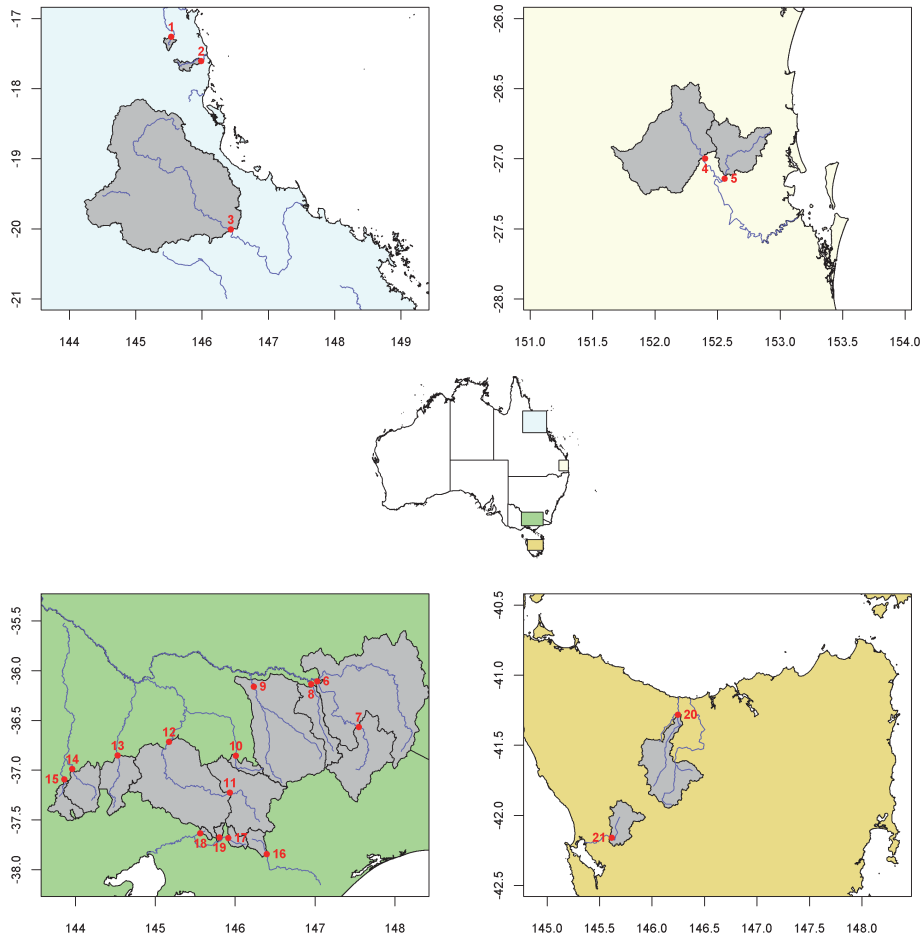
**Fig. 1.** Location of study catchments (location numbers correspond to identifiers in Table 1).

Fig. 2. Skill scores of forecasts made using WAPABA simulations and lag-1 streamflow as predictors plotted against skill scores of forecasts made using selected predictors for the RMSEP (left panel) and CRPS (right panel) skill scores. Each point represents the skill of forecasts for a single location and season. Points above the 1 : 1 line indicate improvements in forecast skill. (Green points are catchments in Queensland, red points are catchments in Tasmania, hollow blue circles tributaries to the upper Murray River, light blue are catchments in Central Victoria and dark blue are catchments in Southern Victoria.)
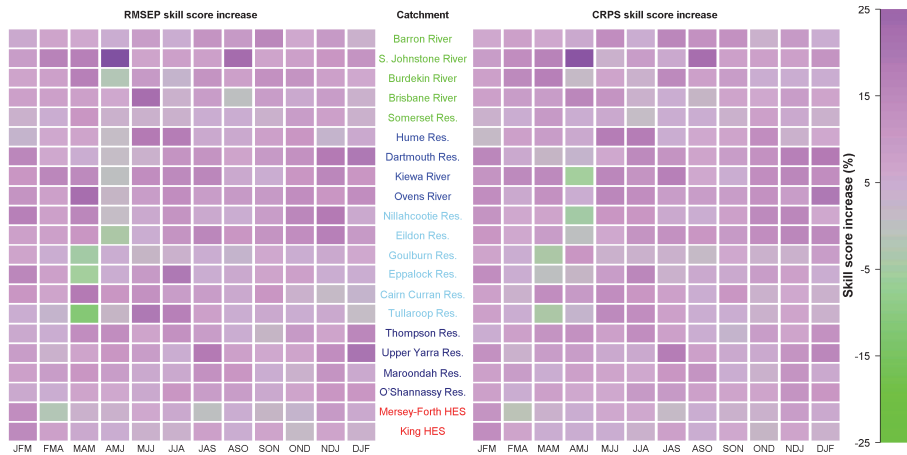
Fig. 3. Increase in skill scores of forecasts achieved by replacing selected predictors representing initial catchment conditions with WAPABA simulations and lag-1 streamflow.
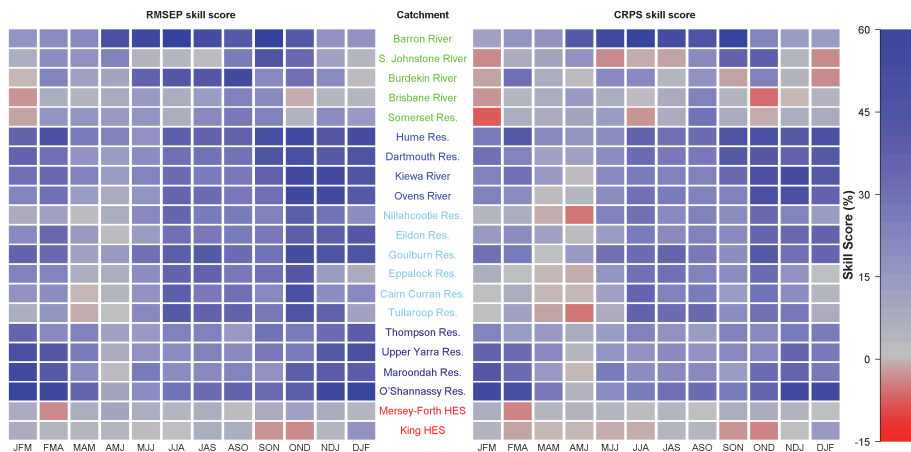
Fig. 4. Skill scores of the cross-validation forecasts made using WAPABA simulations and lag-1 streamflow as predictors.
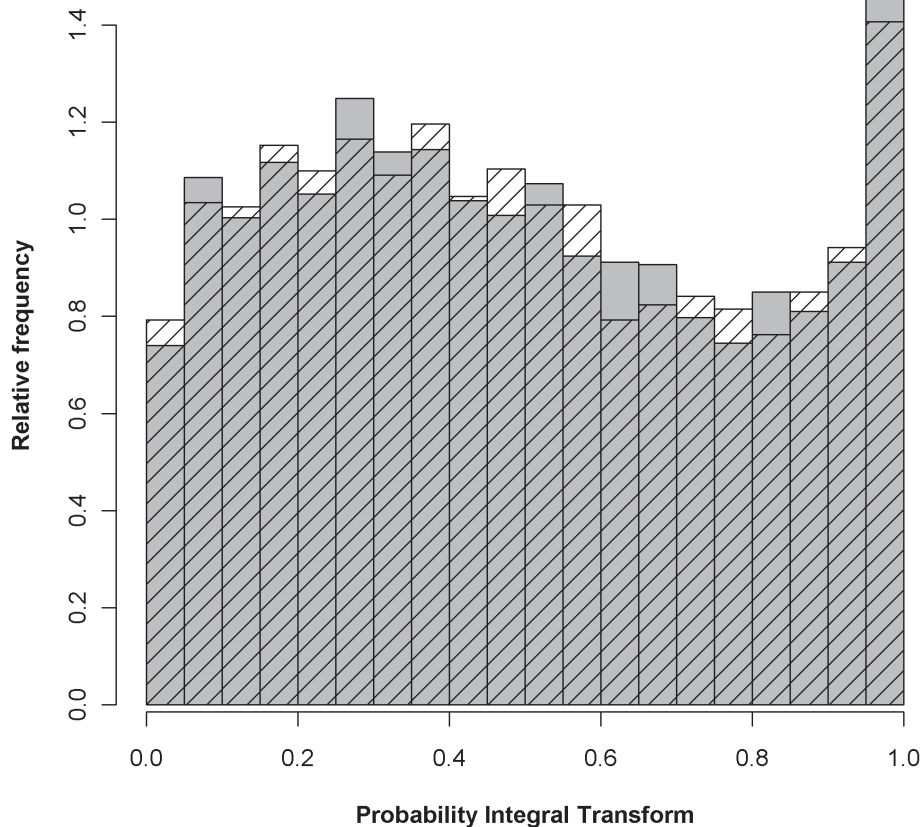
Fig. 5. Probability integral transform histograms illustrating the reliability of forecasts made using selected predictors (solid grey bars) and WAPABA and lag-1 streamflow as predictors (hatched bars).

Fig. 6. Forecasts of Kiewa River inflows into the Murray River for the March-April-May season. Using selected predictors, RMSEP skill score = 9 % and CRPS skill score = 1 %, using WA-PABA simulations and lag-1 streamflow as predictors RMSEP skill score = 18 %, and CRPS skill score = 10 %. (1 : 1 line, forecast median; dark blue vertical line, forecast [0.25, 0.75] quantile range; light and dark blue vertical line, forecast [0.10, 0.90] quantile range; dark grey horizontal line, climatological median; mid gray shade, climatological [0.25, 0.75] quantile range; light and mid gray shade, climatological [0.10, 0.90] quantile range; red dot, observed catchment inflow.)
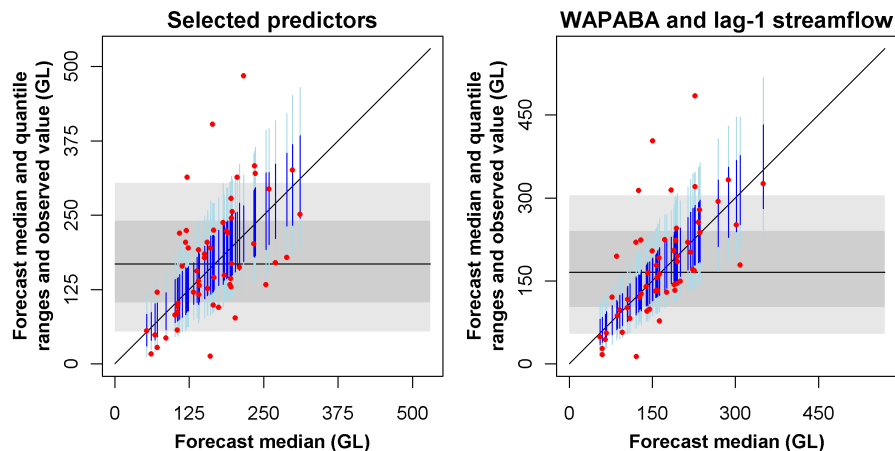
Fig. 7. Forecasts of inflows into Dartmouth Reservoir for the November-December-January season. Using selected predictors, RMSEP skill score = 19 % and CRPS skill score = 17 %, using WAPABA simulations and lag-1 streamflow as predictors RMSEP skill score = 31 %, and CRPS skill score = 28 % (legend as per Fig. 6).
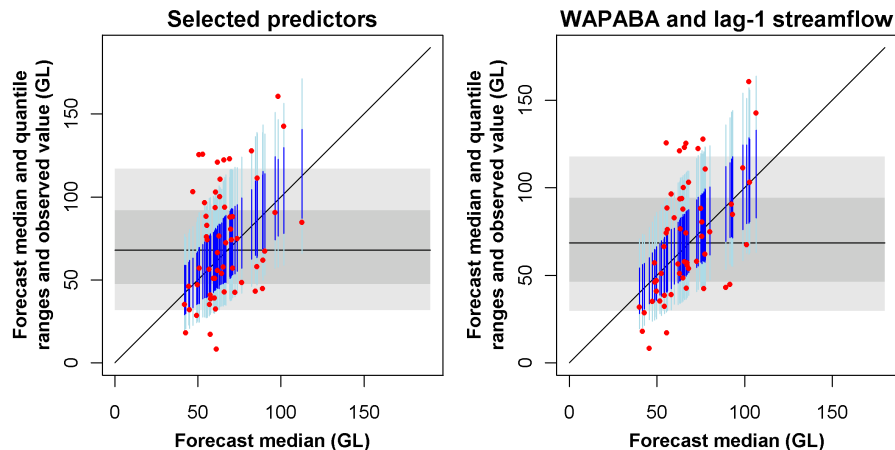
Fig. 8. Forecasts of inflows into Upper Yarra Reservoir for the July-August-September season. Using selected predictors, RMSEP skill score = 5 % and CRPS skill score = 3 %, using WA-PABA simulations and lag-1 streamflow as predictors RMSEP skill score = 17 %, and CRPS skill score = 12 % (legend as per Fig. 6).

Full Screen / Esc

Printer-friendly Version

Interactive Discussion
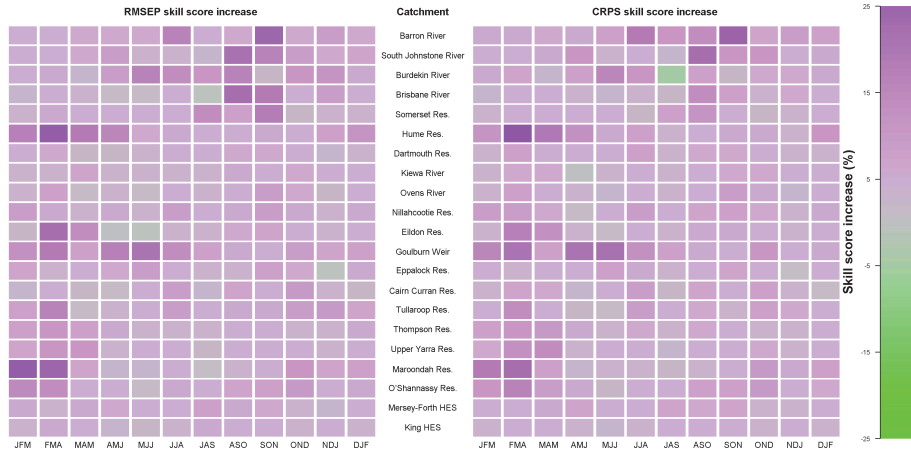
Fig. 9. Increase in skill scores of forecasts achieved by using lag-1 streamflow as well as WAPABA simulations to represent the initial catchment condition.
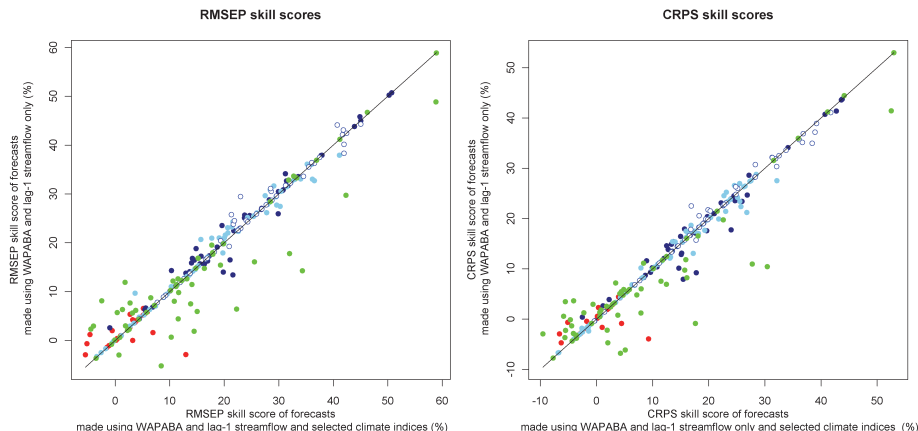
Fig. 10. The contribution of selected climate indices to forecast skill illustrated by plotting skill scores of forecasts made using WAPABA simulations and lag-1 streamflow only as predictors plotted against skill scores of forecasts made using made using WAPABA simulations and lag-1 streamflow and selected climate indices as predictors for the RMSEP (left panel) and CRPS (right panel) skill scores (legend as per Fig. 2).