Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-691, 2017
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 4 January 2017


Hydrology and Earth System Sciences
Discussions

# Value of seasonal streamflow forecasts in emergency response reservoir management

Sean W. D. Turner[1], James Bennett[2], David Robertson[2], Stefano Galelli[3]

[1]SUTD-MIT International Design Centre, Singapore University of Technology and Design, 487372, Singapore.
[2]CSIRO, Melbourne, Clayton, Victoria 3168, Australia.
[3]Pillar of Engineering Systems and Design, Singapore University of Technology and Design, 487372, Singapore.

*Correspondence to*: Stefano Galelli (stefano_galelli@sutd.edu.sg)

**Abstract.** Considerable research effort has recently been directed at improving ensemble seasonal streamflow forecasts, and transferring these methods into operational services. This paper examines the value of forecasts when applied to a range hypothetical reservoirs. We compare forecast-informed reservoir operations with operations based on more traditional control rules established from historical records. Using synthetic forecasts, we show that forecast-informed operations can improve reservoir operations where forecasts are accurate, but that this benefit is far more likely to occur in reservoirs operated for *continually adjusted* objectives (e.g., for hydropower generation) than compared with those operated for *emergency response* objectives (e.g., urban water supply, for which water use restrictions are seldom imposed). We then test whether a modern experimental forecasting system—called Forecast Guided Stochastic Scenarios (FoGSS)—can benefit a wide range of reservoirs operated for emergency response objectives. FoGSS-informed operations improved reservoir operations in a large majority of the reservoirs tested. In the catchments where FoGSS forecasts sometimes failed to improve operations over conventional control rules, we show that this is partly due to less consistently skilful forecasts at the timing during critical decisions are made.

## 1 Introduction

Coupled natural-engineered water resources systems benefit society in a variety of ways. A properly functioning system can ensure reliable public water supply, support agricultural and industrial activity, produce clean hydroelectricity, provide amenity, sustain ecosystems and protect communities against damaging floods. But these benefits are by no means guaranteed; the performance of a system depends largely on the quality of its operating scheme and the intelligence used to support key management decisions. Of particular importance is the operator's ability to inform decisions with accurate estimates of near-term water availability. Flexible, real-time operating schemes that adapt in response to seasonal streamflow forecasts are thus the vanguard of water resources management practice (Rayner et al. 2005, Brown 2010, Gong et al. 2010, Brown et al. 2015). The widely held view is that that real-time (or "online") operating schemes could supplant conventional, "offline" operating schemes, which specify fixed management decisions, or "control rules," as a function of system state variables like time of year, snowpack depth, soil moisture, total water held in reservoirs, and so on.

We suggest two main reasons for recent interest in forecast-informed water resources management. First, it has become apparent that offline control rules can misguide the operator when a system undergoes significant change. A climatic shift, for example, can cause breakdown in established relationships between current system state variables and future water availability, leading to suboptimal operating decisions (Turner and Galelli 2016a). Second, the science of seasonal streamflow forecasting has advanced significantly in recent years. A number of new ensemble seasonal streamflow forecasting methods have been developed, adding to existing ensemble streamflow prediction (ESP) and regression methods (e.g., Wang and Robertson 2011; Olson et al. 2016; Pagano et al. 2014; see review by Yuan et al. 2015). New seasonal streamflow forecast services are becoming available in countries such as the United States, Australia and Sweden.

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-691, 2017
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 4 January 2017


Simultaneously, an emerging field of research has begun to demonstrate the value of seasonal streamflow forecasts when applied to real-world water management problems, such as determining the appropriate water release from a reservoir—the focus of the present study. Water release decisions can be demonstrably improved with seasonal forecasts in a variety of situations, including hydropower dams (Kim and Palmer 1997, Faber and Stedinger 2001, Hamlet et al. 2002, Alemu et al. 2010, Block 2011), water supply reservoirs (Anghileri et al. 2016, Zhao and Zhao 2014, Li et al. 2014) and reservoir systems operated for multiple competing objectives (Graham and Georgakakos 2010, Georgakakos et al. 2012).

Whilst the studies cited above are highly informative, providing strong motivation for further integration of seasonal forecast services into water resources management practice, they tend to focus on systems for which the water release decision is adjusted at frequent intervals—an operating mode we refer to as *continually adjusted operation*. Generally, reservoirs operated for flood control, hydropower and amenity fall into this category, as do low-reliability water supply reservoirs with short critical drawdown periods. This contrasts with *emergency response operations*, for which key management decisions are made relatively infrequently. The most obvious case is a drought action on a large, urban water supply system, such as water use restrictions imposed once every twenty years by design. Emergency response operations may need to be studied as a separate problem to continually adjusted operations, for the following reasons: the stakes tend to be very high and consequently the decisions are often highly politicized, attracting significant public attention (Porter et al. 2015); and the infrequent nature of the decision means that system performance will be determined solely by decisions made at critical moments, such as at the onset of major drought. The ultimate value of a forecast-based operating scheme will depend on the forecast quality during those critical moments, as well as on the operator's willingness to act.

In order to explore the value of applying seasonal forecasts in emergency response operations, we conduct two simulation experiments using reservoir inflow time series recorded at four contrasting catchments located in Australia. The purpose of our first experiment is to determine whether there is any difference between the two operational settings—continually adjusted and emergency response—in terms of the value they reap from seasonal streamflow forecasts. To address this question we compare the relationship between forecast performance and operational value for each case. This is achieved by fabricating for each catchment a hypothetical reservoir and 1000 synthetic inflow forecasts of varying quality, from near-perfect to low-skilled, and then simulating using a rolling horizon, adaptive control operating model. As we shall see, the contrast between the two operational settings is striking. The purpose of our second experiment is to understand whether this result can be explained by forecast performance during periods for which operating decisions become critical. For this experiment we apply an advanced seasonal streamflow forecast system. By varying the reservoir design for each catchment we shift the critical decision points onto different periods, allowing us to investigate the importance of isolated moments of forecast performance to the overall value of the forecast in operation. Results provide insight into the risks operators take when applying a seasonal forecast to critical management decisions in emergency response systems.

## 2 Materials and methods

### 2.1 Inflow records and forecasts

Our experiments are based on four reservoir inflow records (Table 1), which were selected because they represent a range of hydrological regimes (perennial, ephemeral, intermittent) across different regions of Australia. For each inflow record, we study the period 1982 – 2010 (Figure 1), for which forecasts are available.

#### 2.1.1 Synthetic forecasts: Martingale Model of Forecast Evolution (MMFE)

Our first experiment is a sensitivity test for forecast value as a function of forecast quality. To generate many forecasts of varying quality, we use the Martingale Model of Forecast Evolution (MMFE) (Heath and Jackson, 1994). This model can be

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-691, 2017
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 4 January 2017




considered superior to one that simply imposes random error on observed values, since it captures the way in which forecast error decreases as the horizon shortens and more information becomes available to the forecaster (known as the evolution of forecast error) (Zhao et al. 2011). Here we implement a parameter that allows us to control the error of the generated synthetic forecast. This "error injected" parameter takes values between 0 and 1, where 0 implies a perfect forecast and 1 yields a sufficiently error-laden forecast to ensure that our experiments include a wide range of forecast performance. (Note that an error injected of 1 should not be interpreted as having any physical meaning, such as equivalence to climatology.) We generate 1000, 12-month ahead, monthly-resolution synthetic forecasts of varying quality by sampling from a uniform distribution between 0 and 1 to feed the injected error parameter. Each forecast should be considered a separate deterministic forecast rather than a member of a forecast ensemble. Figure 2 displays the goodness-of-fit for these forecasts as a function of the error injected at each forecast lead-time (forecasted against observed values for the period 1982 – 2010). The goodness-of-fit measure is the normalised Root Mean Squared Error (nRMSE), which is the RMSE divided by the standard deviation of observations. Since zero error corresponds to the perfect forecast, all lead times have nRMSE of 0 when no error is injected. As the injected error increases, the performance gap between short and longer lead time forecasts widens, reflecting a deterioration of forecast performance that one would expect with a weaker forecasting system.

#### 2.1.2 Actual forecasts: Forecast Guided Stochastic Scenarios (FoGSS)

In our second experiment, we apply the forecast guided stochastic scenarios (FoGSS) experimental streamflow forecast system. The method behind FoGSS is complex, and accordingly we only give an overview here. A full description, including detailed equations, is available in Bennett et al. (2016) and Schepen and Wang (2014). FoGSS post-processes climate forecasts from the Predictive Ocean and Atmosphere Model for Australia (POAMA) (Hudson et al., 2013; Marshall et al., 2014), with the method of calibration, bridging and merging (CBaM; Schepen and Wang 2014, Schepen et al. 2014, Peng et al. 2014) to produce ensemble precipitation forecasts. CBaM corrects biases, removes noise, downscales forecasts to catchment areas and ensures ensembles are statistically reliable. The precipitation forecasts are then used to force the monthly Water Partitioning and Balance (Wapaba) hydrological model (Wang et al. 2011). Hydrological prediction uncertainty is handled with a 3-stage hydrological error model, which reduces bias and errors, propagates uncertainty, and ensures streamflow forecast ensembles are reliable (Wang et al. 2012; Li et al. 2013; Li et al 2015; Li et al 2016). In months where forecasts are not informative, FoGSS is designed to return a climatological forecast. FoGSS produces 1000-member ensemble streamflow forecasts in the form of monthly-resolution time-series with a 12 month forecast horizon.

FoGSS hindcasts are available for selected Australian catchments for the years 1982-2010 (based on the availability of POAMA reforecasts), including the four catchments examined in this study. The hindcasts are generated using a leave-5-years-out cross-validation scheme (Bennett et al. 2016), which ensures that the performance of FoGSS hindcasts are not artificially inflated. We characterise forecast performance with a skill score calculated from a well-known probabilistic error score, the continuous ranked probability score (CRPS; see, e.g., Gneiting and Raftery, 2007). The skill score is calculated by:

$$CRPSS = \frac{CRPS_{Ref} - CRPS}{CRPS_{Ref}} \times 100\% \qquad \text{Equation 1}$$

where $CRPS$ is the error of FoGSS forecasts and $CRPS_{Ref}$ is the error of a reference forecast, in this case a naïve climatology. The climatology reference forecast is generated from a transformed normal distribution (Wang et al. 2012), fitted to streamflow data using the same leave-5-years out cross-validation as applied to the FoGSS forecasts (Bennett et al. 2016).

FoGSS exhibits a range of performance across the catchments used in this study (Figure 3). In the Upper Yarra, Burrinjuck and Eppalock catchments, FoGSS forecasts are generally skilful at lead times of 0-2 months, extending to more than 3 months at certain times of year (in particular for the Upper Yarra and Burrinjuck catchments). Skill is much less evident in the Serpentine catchment, only appearing evident in a few months of the year (January, August, September, November), even at short lead times. Generally, at longer lead times forecasts are at worst similar to climatology. The only exception is the Eppalock catchment for February and March, where strongly negative skills occur. In the Eppalock catchment, February and March usually experience very low (to zero) inflows. FoGSS forecasts in the Eppalock catchment are slightly positively biased at longer lead times, with these small biases resulting in high relative errors in February and March. However, because inflows are so low during these months, these relative errors have very little influence on annual (or even seasonal) water balances.

## 2.2 Reservoir setup

### 2.2.1 Reservoir model and design specifications

We use monthly resolution reservoir simulation and operating schemes. Each reservoir obeys basic mass balance, meaning volume of water held in storage ($S_{t+1}$) is equal to the previous month's storage ($S_t$) plus total inflow to the reservoir ($Q_t$) minus volume of water released ($R_t$). (Evaporation and other water losses are ignored for simplicity.) The release $R_t$ is constrained physically to a maximum of the available water in storage plus any incoming inflows during period $t$ (Equation 2).

$$S_{t+1} = S_t + Q_t - R_t \qquad \text{Equation 2}$$

$$subject\ to \quad 0 \leq S \leq S_{cap}\,; \;\; 0 \leq R_t \leq \min(S_t + Q_t, R_{max})$$

where $S_{cap}$ is the capacity of the reservoir and $R_{max}$ is the maximum volume of water that can be released during any time period, taken in this study as twice the release target. All excess water is spilled, i.e., $Spill_t$ = max ($S_t + Q_t - R_t - S_{cap}$, 0).

Rather than using the real-world specifications of the four reservoirs corresponding to our inflow records, we vary the size and operation of reservoirs. This approach gives two important advantages. First, it allows us to specify operating objectives relevant to the study question (continuously adjusted operations versus emergency response). Second, it enables us to examine the value of forecasts for reservoirs sensitive to different types of drought, such that overall forecast value becomes dependent on the quality of the forecast at different time periods in the simulation (necessary for experiment 2).

To fabricate these reservoirs we begin by assuming a time-based reliability of 0.95 in all instances. Time-based reliability is the ratio of failure months—months during which the demand for releases cannot be met in full—to the total number of months simulated. A target of 95% can be considered a realistic service standard, since we assume a standard operating policy where reservoirs release demand in full if water is available to do so (this design assumption means that the resultant reservoirs will be very unlikely to empty when operated with more advanced techniques). A constant demand for water is assigned for eight alternative reservoirs by varying a draft ratio (ratio of demand to mean inflow) for values between 0.2 and 0.9 in increments of 0.1. The reservoir capacity required to achieve the target reliability is then determined for each demand using an iterative simulation procedure (storage-yield-reliability analysis). Table 2, which summarises the reservoir designs, shows that the storage must be increased in order to meet greater demand at the target reliability. As demand and storage increase, drift decreases and critical period increases. Critical period gives the time taken for the reservoir to empty under recorded droughts, whilst drift indicates the presence of within-year or over-year behaviour (drift greater than 1 normally indicates within-year reservoirs that refill and spill each year). In other words, as demand is adjusted, the storage dynamics are affected and the reservoir becomes sensitive to different hydrological events. Appendix 1 provides more detailed definitions of the parameters and variables discussed above. All computations are executed using R package *reservoir* (Turner and Galelli 2016b) using observed inflows for the period 1982 – 2011.

### 2.3 Operating schemes

If we allow that the objective of a reservoir can be described adequately by a mathematical function, we can quantify operating performance by imposing penalty costs for deviations from that objective. But to understand the value of a forecast-informed operating model, we also need to compare that performance against a benchmark. We therefore apply two operating schemes in this study: a *benchmark scheme* that ignores forecasts and a *forecast-informed scheme* that makes use of forecasts. Since we are primarily interested in the value added by applying the forecasts to the operation, we must ensure that the performance differences between the two models are attributable to the forecast information rather than conceptual differences in the operating schemes applied. We therefore select two schemes that are conceptually similar (see section 2.3.2), whilst recognising standard, common practice. Our benchmark scheme specifies control rules to govern the operation of a reservoir. Control rules are established by optimising reservoir operations on the basis of historical inflows. Control rules (often termed “release policies”, “hedging rules”, or “rule curves”) are very commonly applied in practice (Loucks et al. 2005), so they provide a realistic benchmark. Our forecast-informed scheme effectively adjusts those control rules in response to new information available through the forecast.

#### 2.3.1 Conventional operating scheme: control rules

The control rules we devise can be thought of as a look-up table that specifies reservoir release as a function of two state variables: volume of water held in storage (discretised uniformly into a manageable number of values) and month-of-year. In practice—and in simulation—the operator simply observes the current reservoir level and then implements release for the time of year as specified by these rules. These rules are designed with respect to the operating objectives and constraints of the system, and can be considered risk-based in the sense that they are conceived to minimise the expected cost of release decisions across the distribution of the inflow for each month. Costs are based on penalties associated with failure to meet the objectives of the reservoir (described in Section 2.4).

The most rigorous way to design such rules is by optimisation. In this study we use Stochastic Dynamic Programming (SDP), which offers four significant advantages. First, SDP handles non-linearity in both the operation of the system and the objective functions. Second, SDP accounts for the effect of uncertainties, in this case stemming from inflows, on system dynamics. Third, SDP finds the optimal operation for a given model of the system (as opposed to other non-linear approaches that approximate the optimal solution). Fourth, SDP returns a cost associated with each combination of state variables, in this case the volume in storage and the month-of-year, known as Bellman’s function. Bellman’s function is useful for the forecast-informed operating scheme introduced in the following section. The inputs to our SDP model are the reservoir specifications, reservoir objective function and inflow time series, which provides inflow distributions for each month. The control rules are optimised by solving a backwards recursive procedure (Bellman 1956, Loucks et al. 2005), which is detailed in Appendix 2. We retrain the control rules for each year of simulation using the same leave-five-years-out cross-validation scheme employed in FoGSS.

#### 2.3.2 Forecast-informed scheme: rolling-horizon, adaptive control

To inform operations with forecasts, we adopt a *rolling-horizon, adaptive control* scheme—also known as Model Predictive Control (Bertsekas 1976). The idea behind this scheme is that the forecast can be used to run short simulations ($t = 1,2,\ldots,H$, where $H=12$ is the forecast length) to evaluate changes in storage that would be experienced under alternative sequences of release decisions. The release decision sequence ($R_1, R_2,\ldots,R_H$) is optimised to minimise the cost over the forecast horizon $H$ plus the cost associated with the resulting storage state:

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-691, 2017
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 4 January 2017


$$\min_{R_{1,2,\ldots,H}} \left\{ \left[ \sum_{t=1}^{H} C_t(R_t, S_t) \right] + X(S_{H+1}) \right\}$$ Equation 3

where $C_t$ is the penalty cost calculated from the reservoir's objective function, and $X(\cdot)$ is a penalty cost function that accounts for the long-term effects of the release decisions being made. The latter helps avoid a short-term, greedy policy that optimises solely for operations in the following *H* months. We set the function $X(\cdot)$ equal to the Bellman's function obtained when designing the control rules, since it contains costs that represent the risk of a given storage level for each month of the year. By using Bellman's function here we effectively append the forecast-informed scheme to the control rules. In essence, this means that the information contained in the forecast is used effectively to adjust the decisions that would be taken using the benchmark scheme—hence our prior statement that the two schemes are conceptually similar.

The optimisation problem is solved at each time step using deterministic dynamic programming, giving the precise optimal release sequence for the forecast horizon ($R_1$, $R_2$,…,$R_H$). The first of these ($R_1$) is implemented in simulation and the remainder are discarded, since the optimisation is repeated on the next time step as a new forecast is issued (hence the term "rolling-horizon", Mayne et al. 2000).

### 2.4 Operating objectives

We test two operating objectives: one that rewards a judicious response to an emergency (*emergency response objective*) response and one that rewards judicious continual adjustments (*continual adjustment objective)*. The emergency response objective encourages full release of water to meet target demand except under drought conditions:

$$C^{emerg} = \sum_{t=1}^{T} [\max(1 - R_t/D, 0)]^2$$ Equation 4

where *D* is the demand. The squared term creates an impetus to cut back the release to reduce the risk of major shortfalls that would occur if the reservoir failed (i.e., becomes fully depleted). Reservoir failure is often associated with highly damaging consequences, such as large water restrictions imposed on households and businesses. Operators therefore tend to hedge against the risk of failure by cutting back the release in small and frequent increments that are, in the long-run, preferable and ultimately less costly than relatively infrequent major shortfalls that would result from total storage depletion (Draper and Lund 2004).

The continual adjustment objective encourages controlled releases to maintain a target storage level, which could represent operation for flood control (e.g., maintain sufficient flood buffer storage), amenity (e.g., avoid unsightly drawdown) or hydropower (maintain high hydraulic head). The objective penalises deviations from a target storage $S^*$, which is set arbitrarily to 75% of total storage capacity in the present study:

$$C^{cont} = \sum_{t=1}^{T} (1 - S_t/S^*)^2$$ Equation 5

where *T* is the final month of the simulation.

Figure 4 gives storage behaviour and release decisions implemented for 0.95 reliability reservoirs (draft ratio = 0.5) operated for the water supply (left-hand panes) and storage level (right-hand panes) objectives described above (here the operation is rolling horizon, adaptive control with a perfect 12-month forecast). The figure contrasts the frequency of decision-making for the two types of operation under study. For emergency response operations we find that the release is adjusted only under drought—predominantly during the Australia's Millennium Drought—and that there are multi-decade periods in which the operator simply releases to meet demand. For the continually adjusted operating setting we find that the release must be adjusted constantly through the operating horizon to keep storage close to the target level of 75%. This contrast shows that the two chosen objectives are suitable for representing emergency response and continually adjusted operation in this study.

## 3 Experiment 1 – Comparing impact of operating mode on value of forecast-informed operations

### 3.1 Experiment description

The purpose of the first experiment is to examine whether a change in operating objective (continually adjusted versus emergency response) affects the relationship between forecast quality and forecast value in operation. For this experiment we hold the reservoir design specifications constant (mid-range draft ratio of 0.5 selected for all four inflow time series). For each reservoir we follow these steps:

1. A set of control rules is optimised with the SDP approach over the period 1982-2010, where the objective is to minimise the sum of penalty costs over the simulation.
2. The adaptive control, rolling horizon scheme is run for a synthetic forecast generated by MMFE over the 1982-2010 period. The value of the forecast is measured by the percentage reduction in penalty cost relative to the control rules over the entire 1982-2010 period.
3. Step 2 is repeated 1000 times, once for each set of synthetic forecasts generated with the MMFE.
4. Steps 1-3 are executed twice—once for emergency response objective and once for the continual adjustment objective. The exact same set of 1000, monthly resolution, 12-month-ahead MMFE forecasts is applied in each case.

We then assess the performance of the forecast-informed operating scheme against the forecast error injected by the MMFE.

### 3.2 Results

Figure 5 shows the value of forecasts achieved by the forecast-informed scheme for each of the four inflow time series. Value of forecasts is presented as the relative reduction of costs (%) with respect to control rules: positive/negative values indicate the forecast-informed scheme outperforms/underperforms the control rules. Forecasts with zero error uniformly outperform control rules, regardless of the objective. As error increases, the value of forecasts behaves differently for the emergency response (a – d) and continually adjusted (e – f) objectives. For the emergency response objective (Figure 5a-d), forecast value becomes relatively unpredictable as soon as error is introduced into the forecast. This contrasts markedly with the continually adjusted objective (Figure 5e–h), for which the forecast value generally decreases relatively slowly as forecast quality is eroded. For errors up to ~0.4 the points are tightly grouped and the forecasts are valuable, showing that forecast error correlates strongly with forecast value for these reservoirs. Taking Burrinjuck (Figure 5a) as an example, we find that an injected forecast error of 0.2 could result in cost reductions anywhere from -5% to +40% for emergency response operations (i.e., the forecast-informed operations could be outperformed by simple control rules by up to 5%). The same forecasts applied to continually adjusted operations (Figure 5e) result in cost reductions in the narrow region of 24 to 26%.

For the emergency response objective, the Burrinjuck and Serpentine reservoirs are particularly sensitive to forecast errors, as errors <0.2 can result in forecast-informed operations being outperformed by control rules. We have numerous forecasts of low error that deliver consistently strong performance in the continually adjusted operational setting, but which can cause a reduction in performance in the emergency response setting. Perhaps surprisingly, we observe several forecasts with high error that deliver consistent reduction in performance in the continually adjusted operational setting, but which can cause improved performance in an emergency operational setting. These results show that the measure of forecast error, quality, skill or goodness-of-fit—if based on the entire forecast period—cannot predict accurately whether that forecast will be valuable in an emergency-type operational setting. This may be because the emergency response objective is constructed to be sensitive to a few serious shortfalls in meeting demand, while the continual adjusted objective rewards consistent performance over all the months assessed. We now turn to experiment 2, which aims at explaining the phenomenon observed here.

## 4 Experiment 2 – The importance of critical drought timing on forecast value

### 4.1 Experiment description

The aim of the second experiment is to determine whether the periods during which critical decisions are made can explain the wide variation in forecast value for a given forecast performance level when applied in emergency response operations. For this experiment we keep the forecast input consistent and instead vary the timing of critical decision points in the simulation. This is achieved by adjusting the reservoir specifications in such a way that they respond to different types of drought (as described in section 2.2.1) so that critical decision periods change. Control rules are designed for all 24 reservoirs (four inflows, eight reservoir set ups) using the SDP approach as above. Operations are then simulated using both the control rules and the forecast-informed model using the median of the value from the full FoGSS forecast ensemble (i.e., the median of the ensemble is taken as the expected inflow at each lead time). While this ignores the spread of the ensemble, we pursue this method because using an ensemble forecast in a multi-stage optimization scheme requires recourse in the control. We cannot simply optimize the release decision by minimizing the expected cost across all ensemble members, because this discounts the operator's ability to adjust the release in response to new information, resulting in over-conservative release decisions and thus weak performance (Raso *et al.*, 2014). The established approach to incorporating information from the spread of the ensemble is Multi-Stage Stochastic Dynamic Programming, which applies a reduced form of the ensemble known as a scenario tree to guide corrective decisions as new forecast data are revealed (Shapiro *et al.*, 2014). Whilst this approach has been applied in a handful of water related studies, including short-horizon problems (Raso *et al.*, 2014) as well using seasonal streamflow forecasts (Housh et al. 2013, Xu et al. 2015), it relies on arbitrary decisions (such as the preferred scenario tree nodal structure), is computationally demanding, and is highly complex, making experimentation laborious and results hard to diagnose. Moreover, our own prior experiments with this approach, in which we reduced the FoGSS ensembles to scenario trees using both the information flow modelling approach (Raso *et al.*, 2013) and the neural gas algorithm (Turner and Galelli 2016c), yielded performances no better on average than those obtained using the median of the ensemble in a deterministic, rolling horizon approach.

We compute performance attained with FoGSS forecasts in relation to the performance attained using a perfect forecast (i.e., a 12-month 'forecast' of the observed inflow):

$$Performance\ Gain = \frac{C^{cntrl} - C^{fcast}}{C^{cntrl} - C^{perfect}} \qquad \text{Equation 6}$$

where $C^{cntrl}$, $C^{fcast}$ and $C^{perfect}$ are the total penalty costs associated with the control rules, forecast-informed operation, and perfect forecast operation respectively. A performance gain of 1 is generally unattainable as it signifies that the forecast is perfect. A performance gain of 0 indicates equal performance with control rules. Negative performance gain suggests that the forecast-based scheme is more costly than control rules.

### 4.2 Results for experiment 2

The left hand panels in Figure 6 (a–d) specify times at which operations become critical (defined as points in which perfect forecast operations implement supply cutbacks). For the reservoirs located in south-eastern Australia (Burrinjuck, Eppalock and Upper Yarra), these periods tend to coincide with the severe Millennium Drought (~2001-2009; van Dijk et al. 2011) occurring towards the end of the simulation horizon. Critical periods for the Serpentine also occur to the end of the record, reflecting long-term declines in inflows over this period (Petrone et al. 2008). For smaller reservoirs, the operations tend to be sensitive to short dry spells, so hedging decisions are required on a more frequent basis during the simulation. The exception is Eppalock, for which the critical period of the reservoir is relatively insensitive to changes in the design demand (Table 2).

The right hand panels (Figure 6e–h) show how operating performance varies with draft ratio. The FoGSS-informed operating model offers performance improvements (i.e., Performance Gain > 0) in more than four fifths of reservoirs tested. Performance gains are achieved for all reservoirs specified for Eppalock and Upper Yarra, and six of the eight reservoirs specified for Burrinjuck. Performance for Serpentine is relatively poor, with only three of seven reservoirs improving under forecast-informed operation (the 90% draft reservoir is omitted in this case, since the reservoir is drawn down at the end of the simulation, meaning the implications of late, sacrificial decisions are unavailable to quantify overall performance). This is partly the result of the generally low skill of FoGSS forecasts with respect to climatology forecasts in the Serpentine catchment (Figure 3), and is also due to the consistency of FoGSS performance through the validation period (discussed in the ensuing paragraphs). Generally, the forecast-informed schemes improved performance over control rules most in reservoirs that must meet high demand (draft ratio > 0.7). For these reservoirs, critical decisions tend to be concentrated in the Millennium Drought period—during which climatology is a poor predictor of inflows, and thus forecast information offers substantial benefits over control rules.

There are certain cases for which seemingly minor changes in the critical decision periods result in large differences in performance gain. To understand this behaviour we can examine specific cases. Figure 8 (a, b) gives storage and release time series (2005–2011) for the Serpentine reservoir with 50% draft requirement. During this period the storage depletes and recovers fully a number of times, and there is a two-year period from mid-2007 to mid-2009 during which storage and inflows are sufficiently healthy that no hedging is required. Performance gain is effectively determined by the differences between control rules and forecast-informed operations during just two periods: the first half of 2007 and the period from mid-2009 to 2011. Overall, the forecast-informed operation improves performance in this reservoir because it instructs the operator to hedge significantly from mid-2009, thus avoiding total reservoir depletion and 100% release shortfall incurred by the control rules. The information provided by FoGSS for this specific time period suffices to avoid reservoir failure and thus reduce the penalty cost by enough to overcome an earlier mistake (the hedge comes too late at the end of 2006). This contrasts with the Serpentine reservoir with 80% draft requirement, for which the forecast causes reduced performance relative to control rules (Figure 8c–d). As for the 50% draft reservoir, we observe an intelligent decision from mid-2009, and the same misstep at the end of 2006. But the 80% draft reservoir never fully recovers after 2006, so all release decisions during this period become locked into memory and contribute to future performance. There appears to be a period in late 2008 during which the forecast performance dips and the operator is instructed to meet the full target release, resulting in costly reservoir failure a few months later. Moreover, the year 2005 also becomes important for this reservoir, and it appears that FoGSS underestimates future flow since an unnecessary and costly hedge is implemented. This simple example demonstrates that a simple shift of emphasis onto some different periods can make the difference between a forecast that outperforms control rules in operation and one that does not. This is consistent with the high sensitivity of the Serpentine Reservoir to forecast error in emergency response operations, demonstrated with the synthetic forecasts in Section 3.2 (this also holds true for the Burrinjuck Reservoir).

We have shown that FoGSS forecasts are skilful, on average, for the 1982-2010 period (Figure 3) but this masks the degree to which skill varies over shorter periods. As we have seen, emergency response objectives can fail in only a few, short periods, meaning that for forecast skill must be consistently available to aid emergency response objectives. To demonstrate the consistency of FoGSS forecast skill, we calculate CRPS skill (equation 1) of lead-0 forecasts for a block of 12 consecutive months, randomly selected from the 1982-2010 validation period. This calculation is repeated by bootstrapping with 5000 repeats. We repeat this procedure for blocks of 2, 3, 4, 5 and 6 years. Figure 7 shows the ranges of skill from the bootstraps as box and whisker plots. The probability that any given 1-year period will have positively skilful forecasts is not statistically significant ($p>0.05$) for all reservoirs. As the blocks get larger, the probability of finding instances of negative skill reduces. For 3-year blocks, forecasts are significantly skilful ($p<0.05$) for both the Eppalock and Upper Yarra reservoirs. However, for Serpentine and Burrinjuck reservoirs, forecasts are not significantly skilful until we test skill for 5-year blocks. That is, FoGSS forecasts are less consistently skilful for the Serpentine and Burrinjuck reservoirs than for the Eppalock and Upper Yarra reservoirs. Less consistent forecast skill helps explain why the forecast-informed scheme does not always outperform control rules in the Serpentine and Burrinjuck reservoirs. An important practical implication of measuring the consistency of skill in this way is that it does not require knowledge of future conditions. This measure can be used to predict the ability of future forecasts to help meet emergency response objectives.

## 5 Discussion and conclusions

We have shown that the benefit to reservoir operators offered by forecasts varies considerably with the objective of the reservoir. For continually adjusted operations, there is a clear relationship between forecast accuracy and benefit: as forecasts become more accurate, operational performance improves. This relationship is much less clear for emergency response objectives, where synthetic experiments showed that even reasonably accurate forecasts may offer little improvement over conventional control rules.

Despite this, we have shown with the real-world example of the FoGSS forecast system that skilful forecasts improve emergency response operation in the majority of reservoirs used in this study, relative to conventional control rules. Meeting the emergency response objective essentially requires effective action in only a few critical instances. Accordingly, we contend that if forecast skill is consistently available, forecasts will better enable emergency response operations objectives to be met. We recommend measuring the consistency of forecast skill as a useful predictor of the value of forecasts to emergency response objectives.

Our results from Experiment 1 show that the operator of an emergency response system will need to accept greater risk than the operator of a continuously adjusted system when adopting a given seasonal forecast service. This may explain the reluctance of operators of large urban water supply systems to adopt seasonal forecasts—an inaccurate forecast at the critical moment may humiliate managers if the implications of missteps are felt by the public. Slow response to an oncoming drought resulting from overestimation of water availability could result in grave consequences in an urban system. For example, the severe rota cuts imposed on millions of people in Sao Paulo have been attributed to tardy management decisions at the onset of a major drought (although in this case the failed management actions were attributed to political factors rather than bad operating decisions) (Meganck et al. 2015). On the other hand, an underestimate of water availability can lead to over-hasty and ultimately unnecessary supply restrictions that may weaken the operator's ability to act decisively the next time a drought emerges. Whilst a skilful forecast service would actually improve these decisions on average over a very long period of time (given enough decision points), managers of such systems may experience only a few such episodes in their entire careers. By adopting a new operating scheme they expose themselves to attack in the event that the scheme fails to work at the time that matters most. Their reluctance to adopt a forecast-informed operating scheme is understandable in this light.

Lastly, the high variability of the performance of emergency response systems present potential pitfalls for case studies assessing the value of forecasts. The weak relationship between forecast accuracy and operating performance means that even good forecasts may result in poor operational performance. The converse may be even worse for operators: by chance, mediocre forecasts may show strong performance for emergency response objectives, giving operators false confidence in the forecasts. For emergency response operations, then, we offer three recommendations:

1. That sensitivity of a given system to forecast performance be assessed, perhaps with synthetic forecasts as in our study;
2. That long records are used to assess performance, and if these are not available that the conclusions of the study be moderated accordingly;
3. That the consistency of forecast skill be established, over the longest period possible, under stringent cross-validation.

The onus must be on the analyst to determine whether the forecast service is sufficiently reliable to satisfy the operator's averseness to adopting a management system that might cause more harm than good during his or her short career.

## 6 Summary

The increasing improvement and availability of seasonal streamflow forecasts opens new opportunities for the adoption of adaptive operating schemes to inform water resources management. Consequently, research investigations need to determine the potential value of forecasts for a range of design and operating settings to which those forecast might be applied. This can be done by measuring system performance improvements as defined by the operating objectives. We used a rolling-horizon, adaptive control approach to demonstrate that the relation between forecast performance and operational value varies significantly when comparing continuously adjusted and emergency response operational settings. We demonstrate a clear and strong relation between forecast skill and value for continuously adjusted operation—operational value increases as the accuracy of the forecast improves. In contrast, good forecast accuracy across the simulation period does not necessarily translate into performance improvement for emergency response systems. We demonstrate with an experimental forecast system, FoGSS, that forecasts can benefit emergency response operations in a number of settings, with several notable exceptions. In all systems, the driver of operating performance is the forecast accuracy at the timing during which a small number of critical decisions are made.

**APPENDIX 1 – Definitions of reservoir parameters and analysis techniques**

All reservoir analyses executed in this study comply with standard, common techniques outlined in mainstream literature (e.g., Loucks et al. 2005, McMahon and Adeloye 2005).

*Time based reliability:*

For a monthly time series, the time-based reliability considers the proportion of months during the simulation period that the target demand is met in full, namely

$$Reliability = \frac{N_s}{Total\ number\ of\ months} \qquad \text{Equation 7}$$

$$0 \leq Reliability \leq 1$$

where $N_s$ is the number of months that the target demand is met in full. Whilst the time-based reliability chosen in this study is 0.95, this does not necessarily mean that reservoir will fail as frequently as once every twenty months. This is because a fail period typically lasts more than a single month. For this reason the time-based reliability is often close to the annual reliability (years in which failure occurs over total number of years simulated).

*Storage-yield-reliability analysis:*

Storage-yield-reliability analysis refers to the procedure used to determine the storage required to meet a demand (or yield) with a specified target, time-based reliability. This is done using an iterative simulation procedure. First, the demand and a trial storage are implemented in the reservoir model. The reservoir is then simulated assuming standard operating policy. The resulting release time series is analysed to determine the time-based reliability of the trial reservoir. The storage is iterated (bi-section method) according to whether the target is missed or exceeded. After a number of iterations, the design converges on the target reliability and storage capacity is attained.

*Critical period:*

The critical period is taken here as the number of month taken for the reservoir to deplete from full to empty (also known as *critical drawdown period*), assuming standard operating policy. The critical period is a function of the demand, storage capacity, and inflow rate during drought. Some reservoirs experience more than one critical period during a simulation. In such cases we take the average of all critical periods.

*Drift:*

Drift (*m*)—also known as *standardised net inflow*—indicates the resilience of a reservoir as well as its tendency for within-year behaviour.

$$m = \frac{1 - DR}{Cv}$$

where *DR* is the draft ratio of the reservoir (demand over mean inflow) and *Cv* is the coefficient of variation of the annualized inflow time series.

**APPENDIX 2 – Reservoir optimization model details**

Control rules (the benchmark scheme) and the rolling horizon, adaptive control (forecast informed scheme) are trained and simulated using the R package *reservoir* (Turner and Galelli 2016b). To develop control rules, the following optimisation problem is solved using a backwards recursive procedure:

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-691, 2017
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 4 January 2017


$$f_t(S_t) = \min_{R_t} \mathop{E}_{Q_t}\{C_t(S_t, Q_t, R_t) + f_{t+1}(S_{t+1})\} \qquad \forall S_t, t \in \{1, \dots, T\}$$ Equation 8

where $f$ is the optimal cost-to-go function (which gives the cost of the optimal decision at time step $t$+1), $C$ is the penalty cost based on deviation from target operation, S is the volume of water in storage, $R$ is the release from storage and $Q$ is the inflow. Storage is discretized into 500 uniform values, meaning the resulting look-up table comprises a 500 × 12 (months) matrix of releases. Release is discretized into 40 uniform values between 0 and $R_{max}$, where $R_{max}$ is twice the demand. Inflow is discretized according to the bounding quantiles of 1.00, 0.95, 0.7125, 0.4750, 0.2375, and 0.00 (as adopted by Stedinger et al. 1984) and likelihood of each flow class is computed for each month using observed inflow data.

For the rolling horizon, adaptive control (or Model Predictive Control) model, the optimisation problem given by Equation 3 is solved at each time step using deterministic dynamic programming.

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-691, 2017
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 4 January 2017


Hydrology and Earth System Sciences
Discussions
Open Access

**Table 1 – Reservoir inflow data; µ and Cv are the mean and coefficient of variation of the annual flow totals respectively.**

| Inflow site | Regime | µ ($Mm^3$) | Cv | Area ($km^2$) | Record | Lat. | Long. | State |
|---|---|---|---|---|---|---|---|---|
| Burrinjuck (BUI) | Perennial | 1252.1 | 0.90 | 1631 | 1900 – 2014 | -35.00 | 148.58 | NSW |
| Lake Eppalock (EPI) | Ephemeral | 166.8 | 0.82 | 1749 | 1900 – 2014 | -36.88 | 144.56 | VIC |
| Serpentine (SEI) | Intermittent | 58.4 | 0.69 | 664 | 1912 – 2014 | -32.40 | 116.10 | WA |
| Upper Yarra (UYI) | Perennial | 153.3 | 0.43 | 337 | 1913 – 2014 | -37.68 | 145.92 | VIC |

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-691, 2017
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 4 January 2017




**Table 2 – Reservoir design specifications and characteristics for 0.95 reliability reservoirs.**

| | Draft ratio | Design Demand [ML/month] | Drift [-] | Design storage [ML] | Storage ratio [years] | Crit. Period [months] |
|---|---|---|---|---|---|---|
| BURRINJUCK | 0.2 | 18,154 | 1.14 | 56,869 | 0.05 | 8 |
| | 0.3 | 27,231 | 1.00 | 144,328 | 0.13 | 11 |
| | 0.4 | 36,308 | 0.85 | 403,911 | 0.37 | 32 |
| | 0.5 | 45,385 | 0.71 | 830,274 | 0.76 | 84 |
| | 0.6 | 54,463 | 0.57 | 1,684,913 | 1.55 | 104 |
| | 0.7 | 63,540 | 0.43 | 2,569,949 | 2.36 | 104 |
| | 0.8 | 72,617 | 0.28 | 3,538,639 | 3.25 | 128 |
| | 0.9 | 81,694 | 0.14 | 4,699,488 | 4.31 | 152 |
| EPPALCOK | 0.2 | 2,398 | 0.88 | 57,560 | 0.40 | 102 |
| | 0.3 | 3,597 | 0.77 | 175,038 | 1.22 | 102 |
| | 0.4 | 4,796 | 0.66 | 288,938 | 2.01 | 102 |
| | 0.5 | 5,995 | 0.55 | 408,827 | 2.84 | 102 |
| | 0.6 | 7,194 | 0.44 | 535,398 | 3.72 | 146 |
| | 0.7 | 8,393 | 0.33 | 710,445 | 4.94 | 146 |
| | 0.8 | 9,592 | 0.22 | 885,493 | 6.15 | 147 |
| | 0.9 | 10,791 | 0.11 | 1,060,540 | 7.37 | 147 |
| SERPENTINE | 0.2 | 507 | 1.50 | 2,028 | 0.07 | 6 |
| | 0.3 | 761 | 1.32 | 4,081 | 0.13 | 11 |
| | 0.4 | 1,014 | 1.13 | 7,199 | 0.24 | 15 |
| | 0.5 | 1,268 | 0.94 | 11,180 | 0.37 | 15 |
| | 0.6 | 1,521 | 0.75 | 14,701 | 0.48 | 15 |
| | 0.7 | 1,775 | 0.56 | 26,940 | 0.89 | 93 |
| | 0.8 | 2,028 | 0.38 | 53,290 | 1.75 | 100 |
| | 0.9 | 2,282 | 0.19 | 88,335 | 2.90 | 112 |
| UPPER YARRA | 0.2 | 2,121 | 1.91 | 2,375 | 0.02 | 3 |
| | 0.3 | 3,182 | 1.67 | 7,217 | 0.06 | 6 |
| | 0.4 | 4,243 | 1.43 | 14,342 | 0.11 | 9 |
| | 0.5 | 5,304 | 1.19 | 25,615 | 0.20 | 13 |
| | 0.6 | 6,364 | 0.96 | 39,065 | 0.31 | 15 |
| | 0.7 | 7,425 | 0.72 | 63,759 | 0.50 | 24 |
| | 0.8 | 8,486 | 0.48 | 139,233 | 1.09 | 142 |
| | 0.9 | 9,546 | 0.24 | 323,047 | 2.54 | 147 |

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-691, 2017
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 4 January 2017


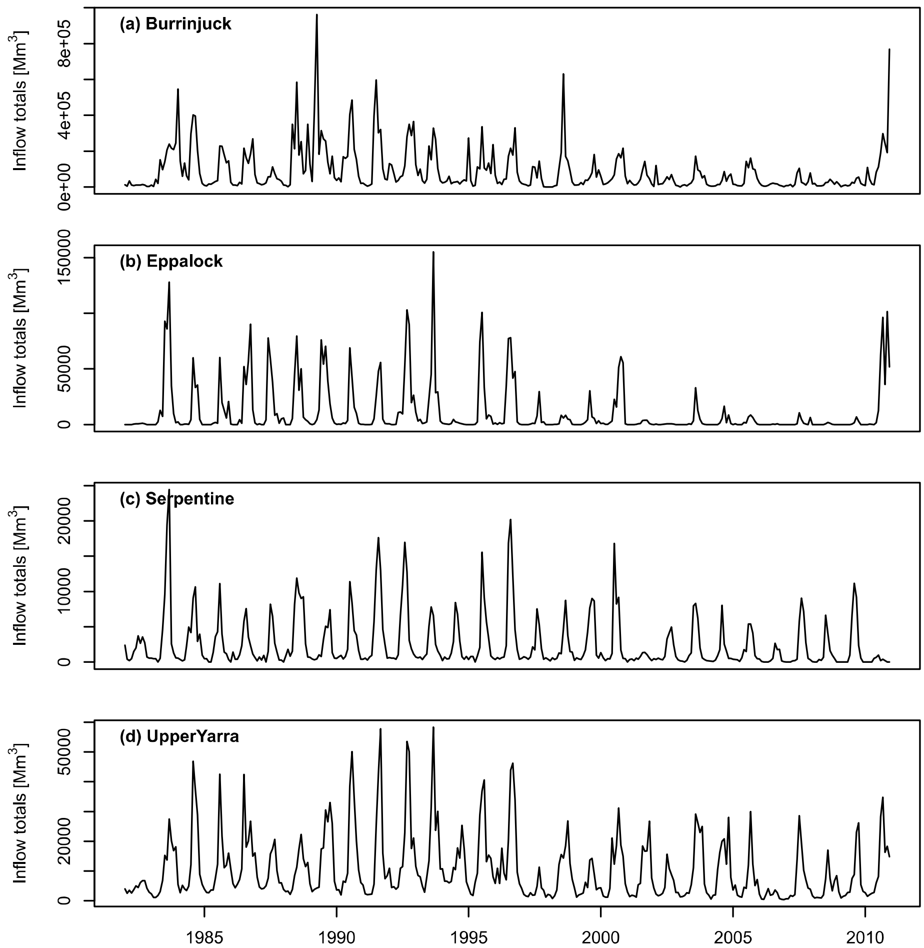


**Figure 1 - Reservoir inflow records for (a) Burrinjuck Dam (BUI), (b) Lake Eppalock (EPI), (c) Serpentine Reservoir (SEI) and (d) Upper Yarra Reservoir (UYI) during the 29-year study period Jan 1982 – Dec 2010.**

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-691, 2017
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 4 January 2017


Hydrology and Earth System Sciences
Discussions

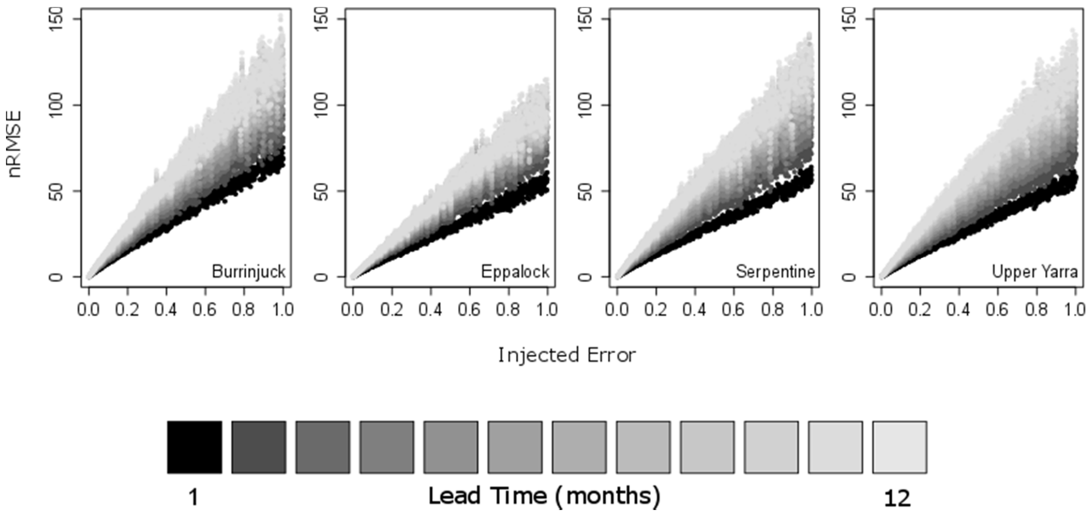


**Figure 2 – Normalised Root Mean Squared Error (nRMSE) for varying error injected into synthetic forecasts generated using the Martingale Model of Forecast Evolution (1000 forecasts, monthly resolution, 12 months ahead, giving 12,000 points on each pane).**

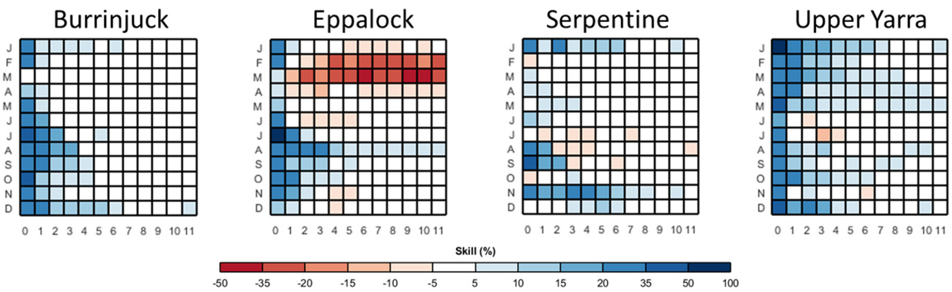


**Figure 3 – Forecast skill measured by the continuous ranked probability score (CRPSS) with respect to climatology forecasts. Rows show target months, columns show lead-time in months.**

Hydrology and Earth System Sciences
Open Access
Discussions

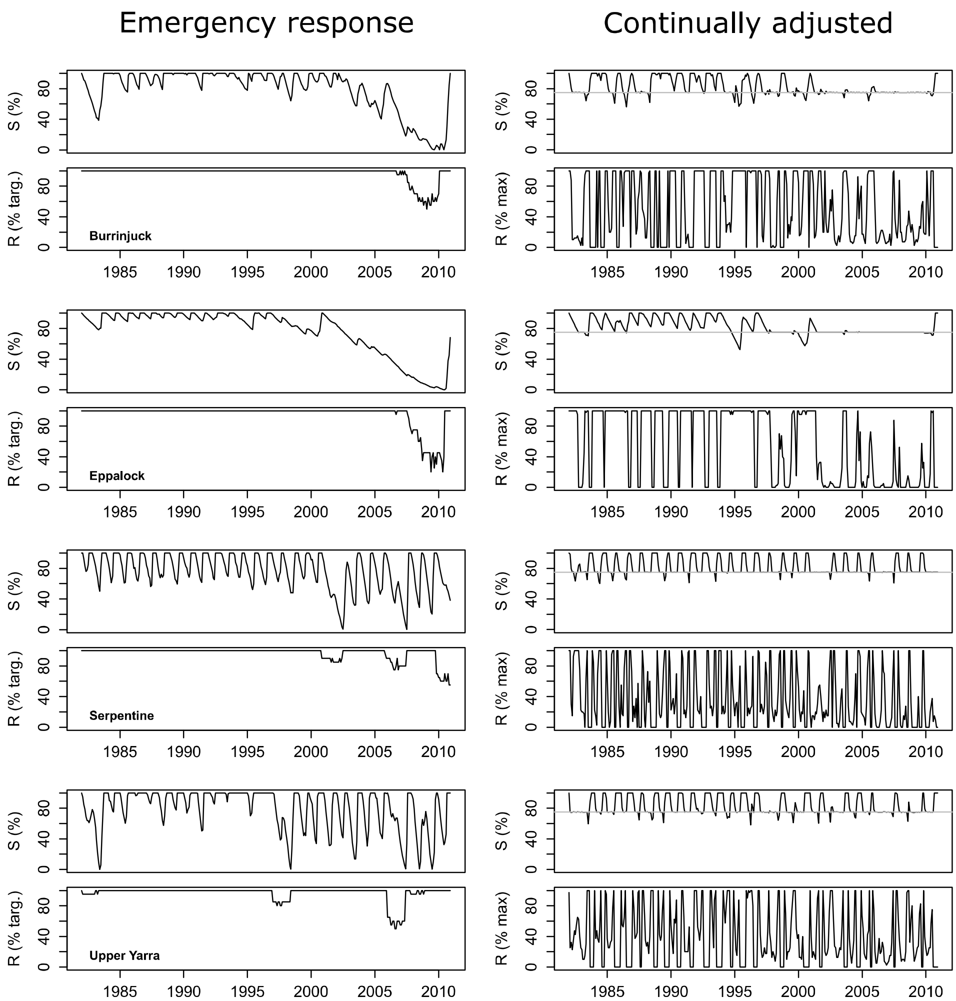


**Figure 4 – Behaviour of reservoirs under emergency response and continually adjusted operations. Simulations use the rolling horizon model with a perfect 12-month (observed) inflow forecast, applied to 95% reliability reservoirs with draft ratio of 0.5. *S* is the storage (as % of capacity) and *R* is the release (given as % of target for emergency response reservoirs and % of maximum possible release for the continually adjusted setting).**

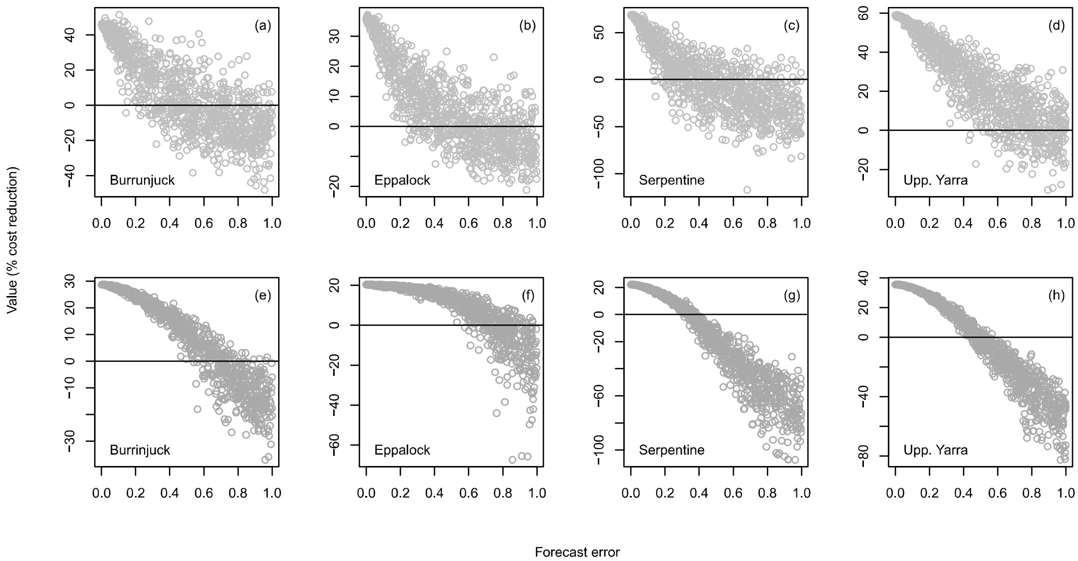


**Figure 5 – Value of the forecast-informed scheme over control rules as a function of forecast error for emergency-response (a – d) and continually adjusted (e – h) operational settings.**

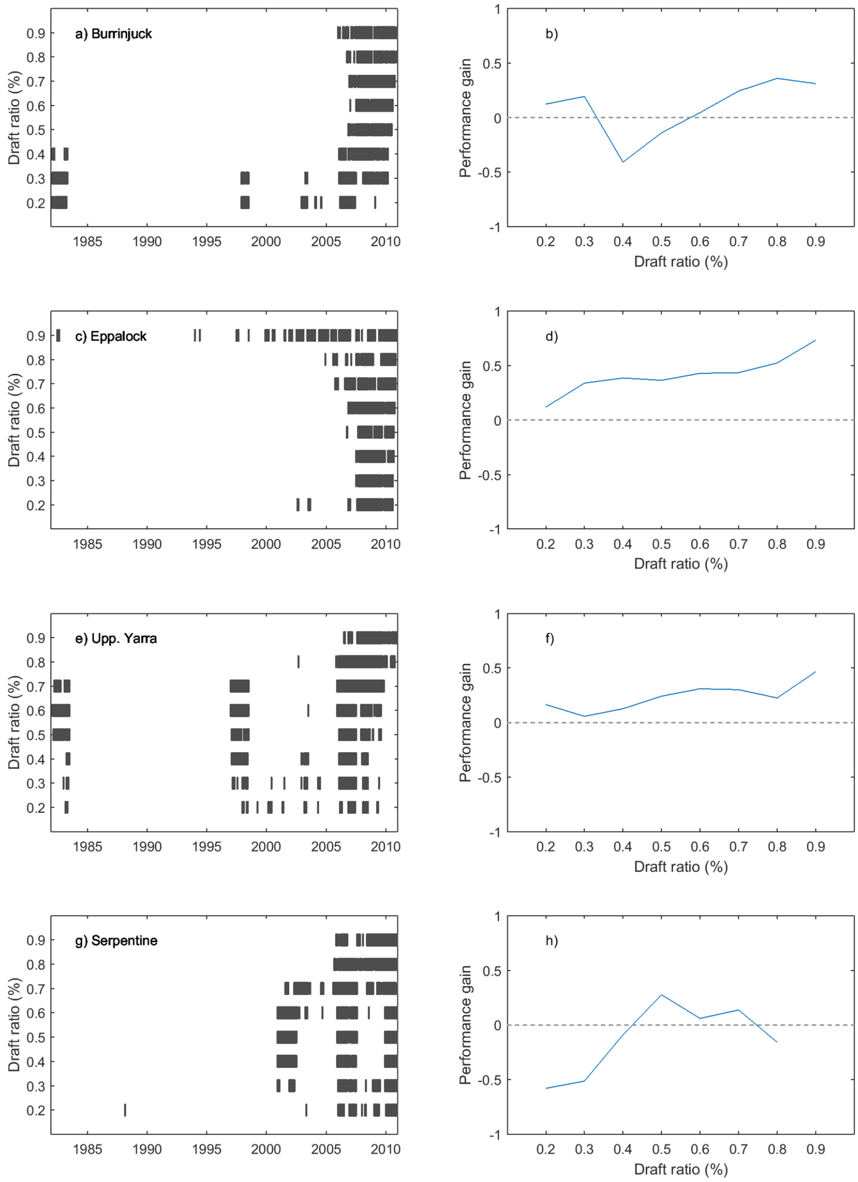


**Figure 6 – Panels a–d give critical decision periods for each reservoir design (draft ratio 0.2, 0.3, …, 0.9). Panels e–h give performance gain plotted against draft ratio.**

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-691, 2017
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 4 January 2017


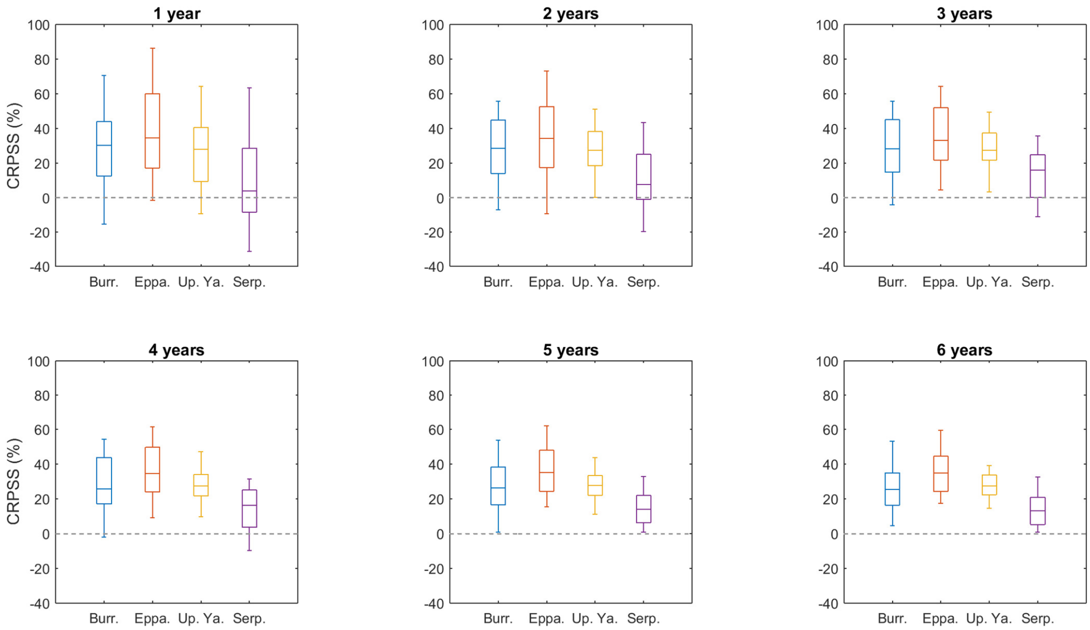


**Figure 7 – Variation in skill lead-0 forecasts for blocks of consecutive months. Skill for consecutive months for blocks of 1-6 years is bootstrapped to create the box and whisker plots. Boxes give interquartile range, whiskers give 90% intervals, lines show median values.**

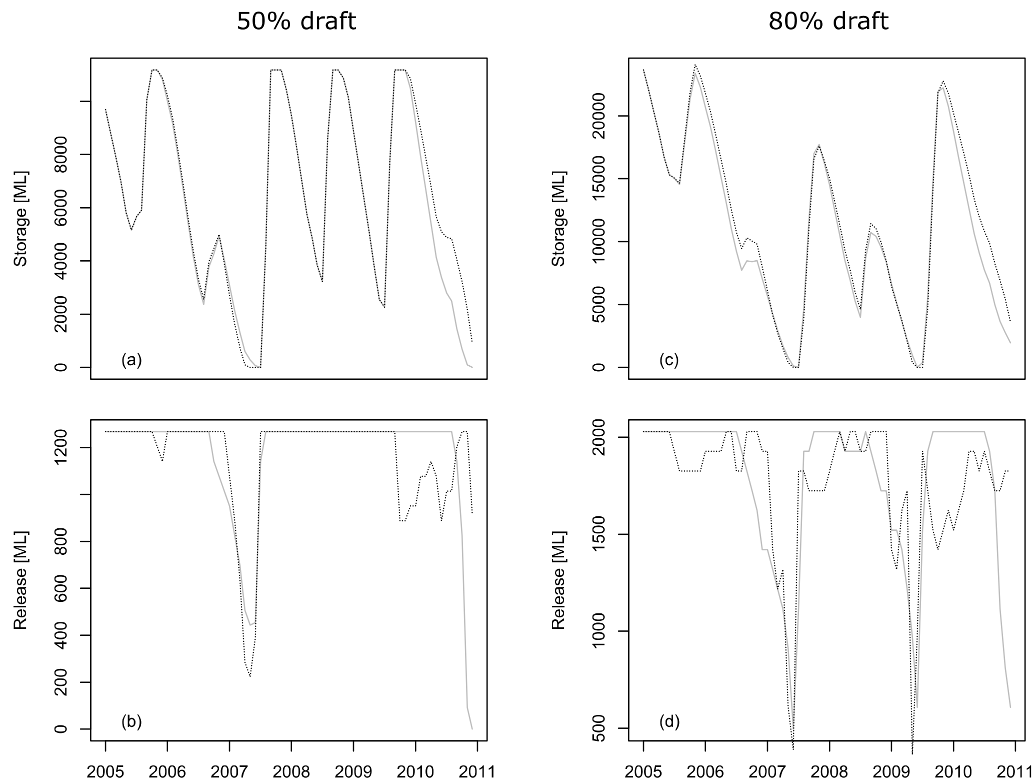


**Figure 8 – Storage and release time series for reservoirs with 50% (a, b) and 80% (c, d) draft ratios. The solid grey line gives operation under control rules whilst the dotted black line gives operation with the FoGSS forecast (median of ensemble).**

5

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-691, 2017
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 4 January 2017




**References**

Alemu, E.T., R.N. Palmer, A. Polebitski, and B. Meaker (2010), Decision support system for optimizing reservoir operations using ensemble streamflow predictions, *Journal of Water Resources Planning and Management*, 137(1), 72-82.

Anghileri, D., N. Voisin, A. Castelletti, F. Pianosi, B. Nijssen, and D.P. Lettenmaier (2016), Value of long□term streamflow forecasts to reservoir operations for water supply in snow□dominated river catchments, *Water Resources Research*, 52(6), 4209-4225.

Bellman, R. (1956), Dynamic programming and Lagrange multipliers, *Proceedings of the National Academy of Sciences of the United States of America*, 42(10), 767-769.

Bennett, J. C., Q. J. Wang, M. Li, D. E. Robertson, and A. Schepen (2016), Reliable long-range ensemble streamflow forecasts by combining dynamical climate forecasts, a conceptual runoff model and a staged error model, *Water Resources Research*, 52(10), 8238-8259.

Bertsekas, D. (1976), Dynamic programming and stochastic control, Academic Press, New York.

Block, P. (2011) Tailoring seasonal climate forecasts for hydropower operations, *Hydrology and Earth System Sciences*, 15, 1355-1368.

Brown, C. (2010), The end of reliability, *Journal of Water Resources Planning and Management*, 136(2), 143-145.

Brown, C.M., J.R. Lund, X. Cai, P.M. Reed, E.A. Zagona, A. Ostfeld, J. Hall, G.W. Characklis, W. Yu, and L. Brekke (2015), The future of water resources systems analysis: Toward a scientific framework for sustainable water management, *Water Resources Research*, 51(8), 6110-6124.

Côté, P., and L. Robert (2016), Comparison of Stochastic Optimization Algorithms for Hydropower Reservoir Operation with Ensemble Streamflow Prediction, *Journal of Water Resources Planning and Management*, 142(2), 04015046.

Draper, A.J., and J.R. Lund (2004), Optimal hedging and carryover storage value, *Journal of Water Resources Planning and Management*, 130(1), 83-87.

Faber, B.A., and J.R. Stedinger (2001), Reservoir optimization using sampling SDP with ensemble streamflow prediction (ESP) forecasts, *Journal of Hydrology*, 249(1), 113-133.

Georgakakos, A.P., H. Yao, M. Kistenmacher, K.P. Georgakakos, N.E. Graham, F.Y. Cheng, C. Spencer, and E. Shamir (2012), Value of adaptive water resources management in Northern California under climatic variability and change: Reservoir management, *Journal of Hydrology*, 412, 34-46.

Gneiting T, Raftery AE. 2007. Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association* **102** 359-378.

Gong, G., L. Wang, L. Condon, A. Shearman, and U. Lall (2010), A simple framework for incorporating seasonal streamflow forecasts into existing water resource management practices, *Journal of the American Water Resources Association*, 46(3), 574-585.

Graham, N.E., and K.P. Georgakakos (2010), Toward understanding the value of climate information for multiobjective reservoir management under present and future climate and demand scenarios, *Journal of Applied Meteorology and Climatology*, 49(4), 557-573.

Hamlet, A.F., D. Huppert, and D.P. Lettenmaier (2002), Economic value of long-lead streamflow forecasts for Columbia River hydropower, *Journal of Water Resources Planning and Management*, 128(2), 91-101.

Heath, D.C. and Jackson, P.L., (1994), Modeling the evolution of demand forecasts ITH application to safety stock analysis in production/distribution systems, *IIE Transactions*, 26(3), 17-30.

Housh, M., A. Ostfeld, U. Shamir (2013), Limited multi-stage stochastic programming for managing water supply systems, *Environmental Modelling & Software*, 41, 53-64.

Hudson, D., A. G. Marshall, Y. Yin, O. Alves, and H. H. Hendon (2013), Improving intraseasonal prediction with a new ensemble generation strategy, *Monthly Weather Reviews*, 141(12), 4429–4449.

Kim, Y.O., and R.N. Palmer (1997), Value of seasonal flow forecasts in Bayesian stochastic programming, *Journal of Water Resources Planning and Management*, 123(6), 327-335.

Li, M., Q. J. Wang, and J. Bennett (2013), Accounting for seasonal dependence in hydrological model errors and prediction uncertainty, *Water Resources Research*, 49, 5913–5929.

Li, W., A. Sankarasubramanian, R.S. Ranjithan, and E.D. Brill (2014), Improved regional water management utilizing climate forecasts: An interbasin transfer model with a risk management framework, *Water Resources Research*, 50(8), 6810-6827.

Li, M., Q. J. Wang, J. C. Bennett, and D. E. Robertson (2015), A strategy to overcome adverse effects of autoregressive updating of streamflow forecasts, *Hydrology and Earth System Sciences*, 19(1), 1-15.

Li, M., Q. J. Wang, J. C. Bennett, and D. E. Robertson (2016), Error reduction and representation in stages (ERRIS) in hydrological modelling for ensemble streamflow forecasting, *Hydrology and Earth System Sciences*, 20, 3561-3579.

Loucks, D.P., E. Van Beek, J.R. Stedinger, J.P.M. Dijkman, and M.T. Villars (2005), Water resources systems planning and management: an introduction to methods, models and applications, Paris: UNESCO.

Marshall, A. G., D. Hudson, M. C. Wheeler, O. Alves, H. H. Hendon, M. J. Pook, and J. S. Risbey (2014), Intra-seasonal drivers of extreme heat over Australia in observations and POAMA-2, *Climate Dynamics*, 43(7), 1915–1937.

Mayne, D., R. Rawlings, C. Rao, and P. Scokaert (2000), Constrained model predictive control: stability and optimality, *Automatica*, 36(6), 789-814.

McMahon, T.A. and A.J. Adeloye (2005), Water resources yield, Water Resources Publications, LLC, Colorado.

Meganck, R., K. Havens, and R. M. Pinto-Coelho (2015), Water: Megacities running dry in Brazil, *Nature*, 521(7552), 289-289.

Olsson, J., C. B. Uvo, K. Foster, and W. Yang (2016), Technical Note: Initial assessment of a multi-method approach to spring-flood forecasting in Sweden, *Hydrology and Earth System Sciences*, 20(2), 659-667.

Pagano, T., A. Wood, K. Werner, and R. Tama-Sweet (2014), Western U.S. Water Supply Forecasting: A Tradition Evolves, *Eos, Transactions American Geophysical Union*, 95(3), 28-29, doi: 10.1002/2014eo030007.

Peng, Z., Q. J. Wang, J. C. Bennett, A. Schepen, F. Pappenberger, P. Pokhrel, and Z. Wang (2014), Statistical Calibration and Bridging of ECMWF System4 Outputs for Forecasting Seasonal Precipitation over China, *Journal of Geophysical Research (Atmospheres)*, 119, 7116–7135.

Petrone KC, Hughes JD, Van Niel TG, and Silberstein RP. (2010), Streamflow decline in southwestern Australia, 1950–2008. *Geophysical Research Letters*, 37(11), doi: 10.1029/2010gl043102.

Porter, M.G., D. Downie, H. Scarborough, O. Sahin, and R.A. Stewart (2015), Drought and Desalination: Melbourne water supply and development choices in the twenty-first century, *Desalination and Water Treatment*, 55(9), 2278-2295.

Raso, L., Giesen, N., Stive, P., Schwanenberg, D., and Overloop, P.J. (2013). Tree structure generation from ensemble forecasts for real time control, *Hydrological Processes*, 27(1), 75-82.

Raso, L., D. Schwanenberg, N.C. van de Giesen, and P.J. van Overloop (2014), Short-term optimal operation of water systems using ensemble forecasts, *Advances in Water Resources*, 71, 200-208.

Rayner, S., D. Lach, and H. Ingram (2005), Weather forecasts are for wimps: why water resource managers do not use climate forecasts, *Climate Change*, 69, 197-227.

Schepen, A., Q. J. Wang, and D. E. Robertson (2014), Seasonal Forecasts of Australian Rainfall through Calibration and Bridging of Coupled GCM Outputs, *Monthly Weather Review*, 142(5), 1758-1770, doi: 10.1175/mwr-d-13-00248.1.

Schepen, A., and Q. Wang (2014), Ensemble forecasts of monthly catchment rainfall out to long lead times by post-processing coupled general circulation model output, *Journal of Hydrology*, 519, 2920–2931.

Shapiro, A., D. Dentcheva, and A. Ruszczynski (2014), Lectures on Stochastic Programming: Modelling and Theory, Vol. 16, SIAM.

Stedinger, J.R., B.F. Sule, and D.P. Loucks (1984), Stochastic dynamic programming models for reservoir operation optimization, *Water resources research*, 20(11), 1499-1505.

Turner, S.W.D., and S. Galelli (2016a), Regime-shifting streamflow processes: Implications for water supply reservoir operations, *Water Resources Research*, 52(5), 3984-4002.

Turner, S.W.D., and S. Galelli (2016b), Water supply sensitivity to climate change: An R package for implementing reservoir storage analysis in global and regional impact studies, *Environmental Modelling & Software*, 76, 13-19.

Turner, S.W.D., and S. Galelli (2016c), scenario: Construct reduced trees with predefined nodal structures, R package version 1.0.0, Comprehensive R Archive Network, Vienna, Austria.

van Dijk, A.I.J.M., H.E. Beck, R.S. Crosbie, R.A.M. Jeu, Y.Y. Liu, G.M. Podger, B. Timbal, and N.R. Viney (2013), The Millennium Drought in southeast Australia (2001-2009): Natural and human causes and implications for water resources, ecosystems, economy, and society, *Water Resources Research*, 49(2), 1040-1057.

Wang, Q. J., and D. E. Robertson (2011), Multisite probabilistic forecasting of seasonal flows for streams with zero value occurrences, *Water Resources Research*, 47, W02546, doi: 10.1029/2010WR009333.

Xu, B., P.A. Zhong, R.C. Zambon, Y. Zhao, and W.W.G. Yeh (2015), Scenario tree reduction in stochastic programming with recourse for hydropower operations, *Water Resources Research*, 51(8), 6359-6380.

Yuan, X., E. F. Wood, and Z. Ma (2015), A review on climate-model-based seasonal hydrologic forecasting: physical understanding and system development, *Wiley Interdisciplinary Reviews: Water*, 2(5), 523-536.

Zhao, T., and J. Zhao (2014), Joint and respective effects of long-and short-term forecast uncertainties on reservoir operations, *Journal of Hydrology*, 517, 83-94.

Zhao, T., X. Cai, and D. Yang (2011), Effect of streamflow forecast uncertainty on real-time reservoir operation, *Advances in Water Resources*, 34(4), 495-504.