PUBLISHED VERSION

Tuteja NK, Shin D, Laugesen R, Khan U, Shao Q, Wang E, Li M, Zheng H, Kuczera G, Kavetski D, Evin G, Thyer M, MacDonald A, Chia T & Le B
Experimental evaluation of the dynamic seasonal streamflow forecasting approach

http://hdl.handle.net/2440/90181

Experimental evaluation of the dynamic seasonal streamflow forecasting approach

Tuteja NK, Shin D, Laugesen R, Khan U, Shao Q, Wang E, Li M, Zheng H, Kuczera G, Kavetski D, Evin G, Thyer M, MacDonald A, Chia T & Le B

Contents

List of tables

List of figures

Acknowledgement

Executive summary

The primary focus of this experimental evaluation project is to answer the key question: *Is it possible to provide accurate and reliable seasonal streamflow forecasts using the dynamic hydrologic modelling approach?* To address this issue, this experimental project evaluated the performance of the dynamic modelling approach for key catchments in the Murray–Darling Basin, where statistical seasonal streamflow forecasts are currently available.

Eight catchments across the southern Murray–Darling basin were examined: (Upper Murrumbidgee) Cotter River at Gingera, Murrumbidgee River above Tantangara Reservoir, Queanbeyan River at Tinderry, and Goobarragandra River at Lacmalac; (Upper Murray) River Murray at Biggara, and Mitta Mitta River at Hinnomunjie; and (Upper Goulburn) Dohertys and Taggerty.

Three conceptual rainfall-runoff models were evaluated: Sacramento, SIMHYD and SMAR. These models were calibrated for 1976–1996 using an objective function that is a combination of the Mean Squared Error (MSE) and the bias in total volume. With the observed rainfall data, Sacramento and SIMHYD showed good performance for most catchments, but late winter and early spring runoff was under predicted in two catchments, mainly due to the absence of snowmelt components in the models.

Forecasting in simulation mode was performed for 1985–2005 with a warm-up period of 1980–1984. Rainfall forecasts from the following two sources were used in the simulations: (a) downscaled POAMA 1.5 rainfall ensemble (ten members and their mean), and (b) historical rainfall ensemble, which is the past ten-year observed rainfall data (ten members and their mean). Forecast skills were assessed using three skill scores: RMSE (Root Mean-Squared Error), RMSEP (Root Mean-Squared Error in Probability) and CRPS (Continuous Ranked Probability Score). NSE (Nash Sutcliffe Efficiency) was also calculated with respect to forecast medians.

POAMA 1.5 rainfall forecasts were downscaled to the catchment scale using a modified version (Shao & Li 2011a) of the analogue method (Timbal, Li & Fernandez 2009). The relationship between POAMA rainfall and AWAP observed rainfall was weak even at monthly aggregated levels. Particularly, rainfall during wet months was frequently underestimated, which considerably reduced the variation range of POAMA rainfall when compared to that of observed data. The serious underestimation of rainfall during wet seasons was a major source of streamflow forecast errors from the rainfall-runoff models.

Two different parameterisation schemes were compared. The general parameterisation scheme optimised one set of SIMHYD parameter values for the entire calibration period, whereas the conditional parameterisation scheme selected 12 optimised sets of SIMHYD parameter values, one set for each month. Since there were marginal differences in skill scores between the two schemes, the general parameterisation scheme was recommended due to its more parsimonious model structure.

Streamflow forecasts driven by downscaled POAMA rainfall forecasts were compared with those driven by historical rainfall data. Compared to the historical rainfall data, POAMA rainfall forecasts produced emphatic streamflows, which were systematically under predicted during wet seasons. In spite of the bias, continued improvements in rainfall forecasts from POAMA and related developments into the future are anticipated to potentially improve seasonal streamflow forecast skills.

ARMA-based posterior bias correction was applied on forecast errors to reduce the systematic bias in streamflow forecasts driven by downscaled POAMA rainfall. The bias correction improved both monthly and three-monthly streamflow forecasts significantly. Bias corrected streamflow forecasts were found to be as accurate and reliable as those from the BJP model. Particularly, during dry seasons, the bias corrected forecasts outperformed BJP in most catchments. The results suggest a very strong case for the use of an ARMA-based bias correction procedure in dynamic seasonal streamflow forecasting and imply the possibility that the current seasonal streamflow forecast service can be upgraded as more robust by blending forecasts from the statistical and dynamic approaches into one forecast product.

Monthly streamflow forecasts from the dynamic modelling approach were found relatively more accurate and reliable compared to three-monthly streamflow forecasts. Considering the value of monthly streamflow forecasts to the Bureau's stakeholders particularly those working on water supply operations, the possibility for a new monthly streamflow forecast service deserves more attention.

As an additional deliverable, a modelling system called the Dynamic Modelling System (DMS) was developed, and its architecture is currently under transition to an operational system.

As a rigorous method for predictive uncertainty estimation, BATEA was applied to target catchments. Using the Bayesian-based Markov Chain Monte Carlo (MCMC) method, BATEA provided more realistic estimates of predictive uncertainty as well as diagnostic evidence for each error source including model structural deficiency and rating curve errors. The results indicated that the technology has a potential to improve not only seasonal streamflow forecast service but also many other services of the Bureau in terms of predictive uncertainty estimation.

In spite of the successful application of dynamic modelling approach in this project, for the national rollout of seasonal streamflow forecast service, further improvements are required. In particular, research efforts are required to improve (a) the accuracy of POAMA rainfall forecasts, (2) rainfall-runoff model performance for dry catchments with intermittent streams, and (3) model performance for catchments with low persistence in their streamflow data.

1. Purpose and motivation

A seasonal climate prediction service has been operating in the Bureau of Meteorology (the Bureau) since 1989 but its primary focus has been on rainfall and temperature rather than water availability. Nevertheless, these monthly updated predictions were highly sought in recent years by Ministers, industry and other government agencies concerned with the possibility of droughts continuing (e.g. Murray–Darling Basin Authority 2009). A seasonal water availability prediction service has been needed in Australia for many years (e.g. Ruiz, Cordery & Sharma 2007) and the Australian Government's recent investment in water information will help address this need (Plummer et al. 2009). Reliable seasonal predictions of streamflows are highly valuable and will have uses for providing water allocation outlooks, informing water markets, planning and managing water use and managing drought (Chiew, Zhou & McMahon 2003).

The opportunity for the Bureau to expand its seasonal prediction service is a result of its new responsibilities, which largely came about because of the impacts of the prolonged drought. In early 2007, the Australian Government announced *Water for the Future*, a $12.9 billion water investment program. This included $450 million for the *Improving Water Information Program* administered by the Bureau and backed by the Commonwealth *Water Act 2007* and key stakeholders. A series of stakeholder workshops across the water sector in Australia were led by the Extended Hydrological Prediction (EHP) Section of the Water Forecasting Branch in 2009. These workshops, which were organised in 2009 against the background of prolonged drought and record low water availability, established a clear need for the Bureau to develop a seasonal streamflow forecasting service.

Seasonal forecasts of water availability can be made using dynamical, statistical or hybrid modelling approaches. The statistical approach is based upon direct relationships derived from observed data and derived predictor indices. Many climate indicators based on atmospheric pressure and sea surface temperature (SST) anomalies have been linked to future seasonal rainfalls. Such relationships were exploited to predict streamflow several months or seasons ahead (e.g. Chiew, Zhou & McMahon 2003; Day 1985; Hammer, Nicholls & Mitchell 2000; Moore, Jones & Black 1989; Plummer et al. 2009). In Australia, statistical prediction methods for water availability are more advanced in terms of their stage of development. The Queensland Department of Primary Industries and Fisheries developed the Rainman decision support system, which provides the probability of flow, based on phases of both the Southern Oscillation Index (SOI) and SSTs. The forecasts are considered skilful out to only one season due to the limited number of predictors (Clewett et al. 2003).

The Bureau launched a new Seasonal Streamflow Forecasting (SSF) service in December 2010. The service delivers three-month ahead probabilistic forecasts of total streamflow volumes at a site or total inflows into major water supply storages (www.bom.gov.au/water/ssf). Forecasts for this service are being derived from a statistical modelling approach called the Bayesian Joint Probability model (BJP; Wang & Robertson 2011; Wang, Robertson & Chiew 2009), which was recently developed within the auspices of the Water Information Research and Development Alliance Program (WIRADA) with CSIRO. The Bureau operationalised BJP for the SSF service (Peaty & Seasonal Streamflow Forecasting Team 2009) by developing a new modelling system called WAFARi (Water Availability Forecasts of Australian Rivers; Shin et al. 2011). The BJP approach transforms a set of streamflows and their predictors into a multivariate normal distribution and infers the distribution of model parameters using a Bayesian formulation, which is implemented through a Markov Chain Monte Carlo (MCMC) sampling method. Currently, the service updates streamflow forecasts every month for 21 key water supply catchments in the Murray–Darling Basin, with the objective of helping water suppliers and users make better decisions on available water resources.

The main focus of this experimental project is to answer the key question: Is *it possible to provide accurate and reliable seasonal streamflow forecasts using the dynamic hydrologic modelling approach?* To address this issue, this experimental project is focused on a rigorous evaluation of the performance of the dynamic modelling approach for key catchments in the Murray–Darling Basin where statistical seasonal streamflow forecasts are currently available (October 2009 – April 2011). This report describes detailed outcomes from the experimental evaluation project. The report is organised as follows: project overview and linkages are listed in section 2; target catchments and climate and hydrology data used in this evaluation are discussed in section 3; section 4 describes the hydrologic models and the modelling approach; the experimental modelling system developed in this project is discussed in section 5; modelling results and evaluation of the seasonal forecasting capability is detailed in section 6; uncertainty analysis results from the work on rigorous treatment of predictive uncertainty in the context of seasonal streamflow forecasting are summarised in section 7; major outputs and key deliverables from this work are discussed in section 8 and recommendations on research-to-operation issues, further research and implementation issues are detailed in section 9.

2. Project overview and linkages

The project team at the Bureau leveraged support through the following collaborations and key initiatives:

- research support from CSIRO on hydrologic modelling and seasonal rainfall downscaling under WIRADA Project 4.2 'Water Forecasting and Prediction – Seasonal and Long-term Forecasting'

- research support from CAWCR on provision of the seasonal climate forecasts derived from POAMA 1.5 as well as support on the technology for rainfall downscaling, which was improved further through the WIRADA initiative

- research support from the University of Newcastle (UoN) and the University of Adelaide (UoA) through the 'BATEA Stage 1 Project' on advanced Bayesian approaches to catchment hydrologic modelling and rigorous treatment of predictive uncertainty

- support from eWater CRC (www.toolkit.net.au) on provision of the catchment hydrologic models used in the experimental Dynamic Modelling System (DMS) developed in this project

- support from the National Climate Centre, Climate Information Services Branch of the Bureau and CSIRO for observed climate data (AWAP [Australian Water Availability Project] gridded rainfall and potential evapotranspiration) and Water Data Branch and a number of key agencies including the Murray–Darling Basin Authority (MDBA) and ACTEW Cooperation Ltd, for provision of the streamflow data.

An overview of the project structure is illustrated in Figure 1 (see following sections for detailed description and findings).

1. First, historical modelling is performed on the target catchments to establish a retrospective water balance through the use of lumped rainfall-runoff models. Each model was driven by observed climate data from AWAP and calibrated against observed daily streamflow data. For each model, five sets of optimal parameter values (i.e. five hydrologic model ensembles) were chosen and used to produce simulated streamflow. Hereafter, streamflow simulated with observed climatology is referred to as 'reference flows'.

2. The rainfall forecast ensembles (ten members) from the Bureau's global climate model POAMA 1.5 (Alves et al. 2003) are then downscaled through an improvisation for bias correction by Shao and Li (2009; 2010; 2011a; 2011b) and Shao et al. 2009. The analogue downscaling method of Shao and Li is a modified version of the original method of Timbal and McAvaney (2001) and Timbal, Li and Fernandez (2009). Statistical downscaling models (SDMs) are mostly based on the view that the regional climate is conditioned by two factors: (a) the large scale climatic state and (b) local physiographic features. A SDM based on an analogue approach was developed within the Australian Bureau of Meteorology and applied to ten regions covering the entire Australian continent. This method searches a past date with the most similar weather pattern as the target date in future and use the predictand, here rainfall, of the past date as the forecast of the target date.

3. Forecasting process outlined above in simulation mode leads to 55 streamflow ensemble time series at daily time steps (five hydrologic models X 11 downscaled rainfall ensembles comprising ten POAMA ensembles and the ensemble mean), which is then aggregated to monthly and three-monthly time steps for verification of the streamflow forecasts using a range of forecast verification skill scores.

4. The difference (or the bias) between observed streamflow and the forecast streamflow results from inaccuracies and collective uncertainties in input rainfall and output runoff data, hydrologic model structure and the use of rainfall forecasts. However, the difference (bias) between the reference streamflow and the forecast streamflow is a direct result of replacing observed rainfall with rainfall forecasts. The biased streamflow is corrected through the use of a multiplicative ARMA model (Box & Jenkins 1976; Hyndman 2011; Tuteja & Cunnane 1999) and improvements thereof in streamflow forecasts are quantified again for forecast verification.

5. Further, full treatment of predictive uncertainty is handled through the use of Bayesian Total Error Analysis (BATEA) technology. Note that a separate detailed report on the BATEA Stage 1 project was prepared by the UoN and UoA and only main findings from the BATEA Stage 1 project and implications for future work are included in this report.

It is expected that the forecasts from dynamic and the statistical approaches can be blended into a hybrid forecast, once both systems have demonstrated a satisfactory level of accuracy and reliability. The exact approach for combining these systems will require further investigation, which is outside the scope of this experimental evaluation project. However, the evaluation results reported here will provide valuable information for these decisions.
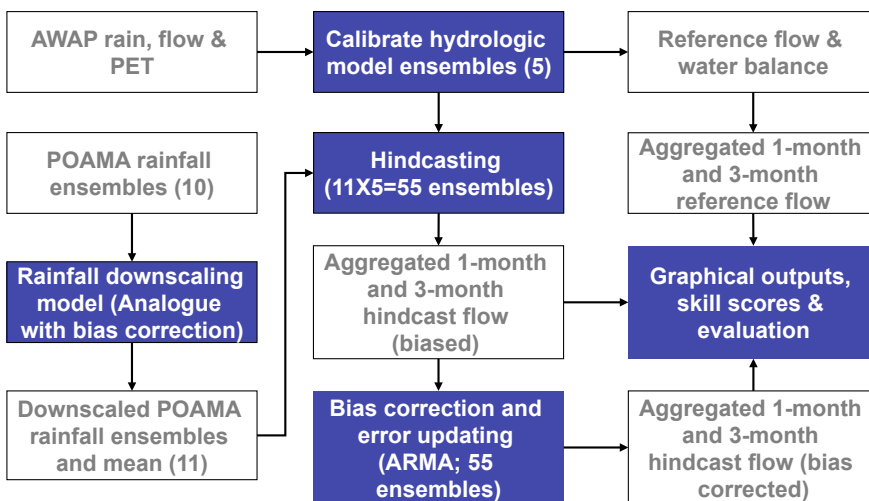
Figure 1: An overview of the project structure

3. Experimental catchments and data

3.1 Selection criteria for the experimental catchments

The target catchments were selected by the following criteria:

- sites must be common to those included in the current Seasonal Streamflow Forecasting service where forecasts are available from the BJP modelling approach

- sites must not have unnatural flows (e.g. downstream of a gated storage)

- sites must be at locations strategically important for water supply (e.g. inflows into a major water supply reservoir)

- data must be concurrent across all sites at some point in time

- at least 30-year streamflow data are available

- catchment climatology and hydrology must be well understood.

On the basis of the criteria, eight target catchments were selected across the Upper Murrumbidgee, Upper Murray and Upper Goulburn valleys in the Murray–Darling Basin in south-eastern Australia for this evaluation project (Figure 2; Table 1).

Figure 2: Locations of the study catchments

Table 1: Locations and characteristics of the study catchments

Station ID	Station Name	Location		Area	Rainfall (1975–2006)	Runoff (1975–2006)	Runoff Coeff.
		Lat.	Long.	(km2)	(mm/yr)	(mm/yr)	(unitless)
Upper Murrumbidgee (Upstream of Burrinjuck Dam)							
410730	Cotter River at Gingera	35.59	148.82	148	1,116	276	0.25
410535	Murrumbidgee River at above Tantangara Reservoir	35.79	148.66	216	1,273*	683	0.54*
410734	Queanbeyan River at Tinderry	35.62	149.35	490	799	124	0.16
410057	Goobarragandra River at Lacmalac	35.32	148.35	673	1,155	384	0.33
Upper Murray							
401012	Murray River at Biggara	36.32	148.05	1,165	1,103	377	0.34
401203	Mitta Mitta River at Hinnomunjie	36.94	147.61	1,533	1,260	258	0.20
Goulburn							
405219	Goulburn at Dohertys (u/s of Lake Eildon)	37.33	146.13	694	1,181	436	0.37
405209	Goulburn at Taggerty (d/s of Lake Eildon)	37.32	145.72	619	1,272	449	0.35

The rainfall average derived from AWAP grids is expected to be underestimated because of the sparse rainfall gauges and alpine terrain of the catchment.

3.2   Catchments

3.2.1   The Upper Murrumbidgee catchment

The Murrumbidgee catchment is bounded by Cooma in the east, Balranald in the west, Temora to the north and Henty to the south. The catchment covers an area of about 84,000 km². The Murrumbidgee River is the main stream running through the Murrumbidgee catchment. The Upper Murrumbidgee catchment (UMC) is bounded to the south and east by the Great Dividing Range, and it forms the south-eastern edge of the Murray–Darling Basin (MDB). It extends from headwaters of the Murrumbidgee River above Tantangara Dam in Kosciuszko National Park, to Burrinjuck Dam – a total of 14,060 km². The UMC is the source of water for communities living in Cooma, Queanbeyan, Canberra and Yass regions, as well as major downstream communities. The main tributaries include the Cotter, Queanbeyan, Goobragandra, Numeralla, Bredbo, Molonglo, Yass and Goodradigbee rivers. Significant sections of these rivers are deeply incised, while there are other areas with relatively narrow and flat floodplains.

The UMC is a major water supply catchment of the MDB. Most of the usable groundwater occurs in the fractured rock aquifers, which cover a large proportion of the UMC. Limited groundwater availability has meant a heavy reliance on surface water in the catchment, with increasing demand from all sectors (www.murrumbidgee.cma.nsw. gov.au). Groundwater resources have not been assessed very well and in areas where groundwater is accessible, this usually leads to groundwater use without consideration being given to local or regional impacts. More than 50% of the UMC is comprised of sedimentary duplex soils, located in semi-arable to non-arable soil landscapes. About 30% of the UMC has granite soil landscapes and about 5–10% are basalt landscapes. High rainfall in the UMC, together with granite and sedimentary duplex soil landscapes, and relatively unaltered land use across large upland areas of the UMC substantially influence the prevailing hydrologic processes; i.e. high runoff coefficient, short flow paths and response times of the underlying groundwater flow systems and high incidence of the saturated excess surface runoff conditions through duplex and granitic soil landscapes.

3.2.2   The Upper Murray catchment

The Upper Murray River Basin forms the catchment of Hume Reservoir and covers an area of 15,280 km^2 in New South Wales and Victoria (Goulburn-Murray Water: www.g-mwater.com.au). The River Murray forms the north and east sides of the Victorian part of the Basin. The Victorian portion of the Upper Murray River Basin covers an area of 10,150 km^2. It is subdivided into two components – the Mitta Mitta River catchment and northern section of the basin which drains into the Victorian bank of the River Murray. Flows are regulated by Hume and Dartmouth reservoirs, owned by Goulburn-Murray Water (G-MW) and operated by Murray–Darling Basin Authority under the Murray–Darling Basin Agreement. The Upper Murray River Basin gains approximately 600 GL per year from inter-basin transfers via the Snowy Mountains Hydroelectric Scheme to the New South Wales section of the Basin.

All major valleys which lie in the north have been cleared for agriculture, although approximately 80% of the Upper Murray Basin still remains forested. The principal forms of land use are water conservation, forestry, grazing and agriculture. Water use accounts for only a very small fraction of the highly regulated resources originating in the Basin. Those resources are committed to water supply requirements throughout the length of the River Murray.

The main source of water is the rain although some precipitation falls as snow above 750 m to 800 m. The land above 1,050 m comprises only 23% of the catchment area and yet produces about 43% of the total yield. Along the Mitta Mitta River, mean annual flow can triple from Hinnomunjie in the south to the north. Reasonably highest flows in October are attributable to the spring snowmelt (Figure 3). In the Upper Murray Basin, the shallow aquifer system consists predominately of outcropping Palaeozoic-aged sedimentary rocks intruded in places by granites which are surrounded by associated metamorphic rocks. Older volcanic basalts cover a small area in the south of the Basin. Many of the streams have significant alluvial deposits along the streamlines. The bulk of the groundwater resource is fresh although a small region containing groundwater of marginal quality is located in the southeast corner of the Basin.
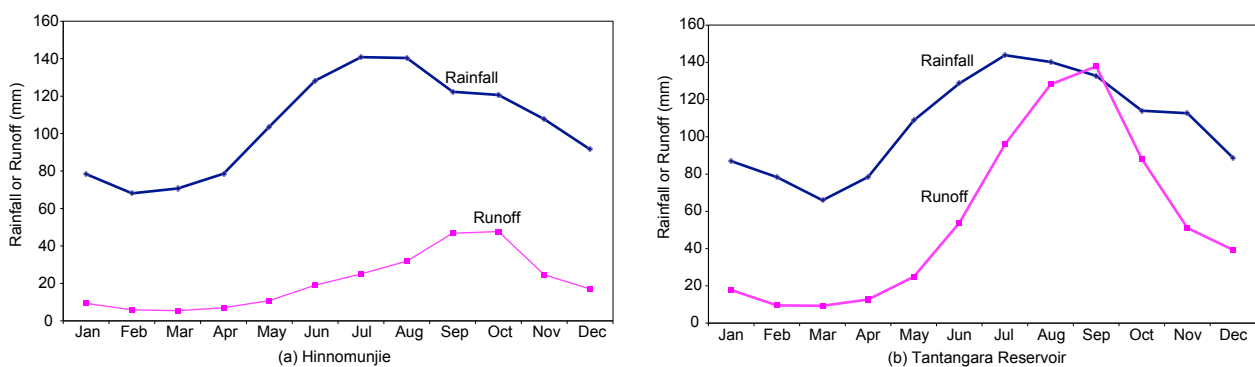
Figure 3: Long-term average of monthly water budget of Hinnomunjie and Tantangara Reservoir to show the impact of snowmelt. The averages were calculated from data between 1978 and 2008

3.2.3   The Goulburn catchment

The Goulburn River Basin covers 16,192 km^2 in central Victoria and extends from the Great Dividing Range near Woods Point, to the River Murray in the northwest near Echuca. The southern boundary of the Goulburn catchment runs along the Hume Range and the Great Dividing Range, and then turns northward to Mt Buller at 1,804 m in the southeast (Goulburn-Murray Water: www.g-mwater.com.au). Lake Eildon lies in the southeast of the Basin and collects flows from the Upper Goulburn, and a number of creeks. Streamflow along the Goulburn River has been modified by two major features, Lake Eildon and the Goulburn Weir.

The three months of greatest flow are July to September, accounting for 52% of the annual flow, and the three months of least flow are January to March, accounting for 5% (Figure 4). Operation of the Eildon Reservoir has reduced the July to September flows passing Eildon to 33% of the annual total, allowing an increase of the January to March flows of 23% of the annual total. The shallow aquifer systems of the Goulburn River Basin occur in three main hydrogeological conditions. The Shepparton Formation aquifer which lies throughout the northern sections of the Basin and is composed of shoestring sands, amongst silt and clay containing brackish groundwater; the remaining central and southern portion of the Basin comprises outcropping basement rocks overlain, in the valleys within the highlands, by Quaternary alluvial sand and gravel. Groundwater in these two units is generally of good quality.

Agriculture in the Basin is diverse, ranging from hardwood timber production in the southeast to dairying and fruit production in the north. The Lake Eildon environs produce sheep for wool, and beef and dairy cattle. Further along the Goulburn Valley, sheep and cropping are important in dryland and irrigated areas.
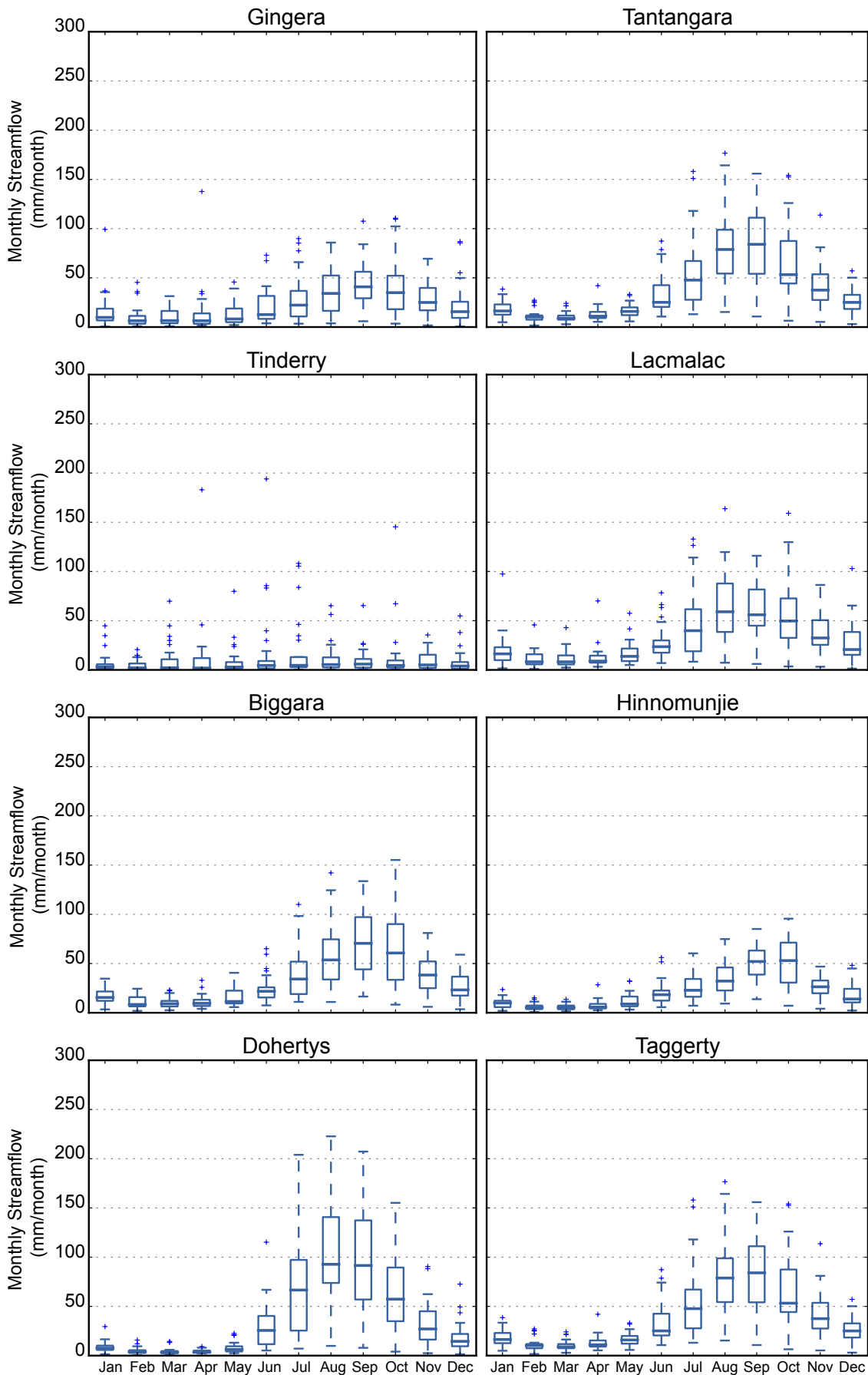
Figure 4: Monthly streamflow of the study catchments

3.3   Observation data

Daily rainfall, areal potential evapotranspiration and streamflow data from 1975 to 2006 were used in this study for calibration and validation of the rainfall-runoff models (Figure 5). Catchment average rainfall data were derived from AWAP (Australian Water Availability Project) grid data with about 5 km spatial resolution (Raupach et al. 2009; Raupach et al. 2011). Monthly rainfalls of the catchments between 1975 and 2006 are shown in Figure 6. Tinderry received noticeably lower rainfall compared to other catchments, because of local orographic effect.

Monthly potential evapotranspiration (PET) data of each catchment were derived from the AWAP modelling results using Priestley Taylor's method (Raupach et al. 2009). Monthly PET data between 1975 and 2005 are shown in Figure 7.

Daily streamflow data at the outlet station of each catchment were acquired from external agencies, including Murray–Darling Basin Authority, ActewAGL, Melbourne Water and Goulburn-Murray Water. Monthly runoff data between 1975 and 2005 are shown in Figure 4. Note that streamflow data for most gauging stations will be available from the Australian Water Resources Information System (AWRIS) in the future.
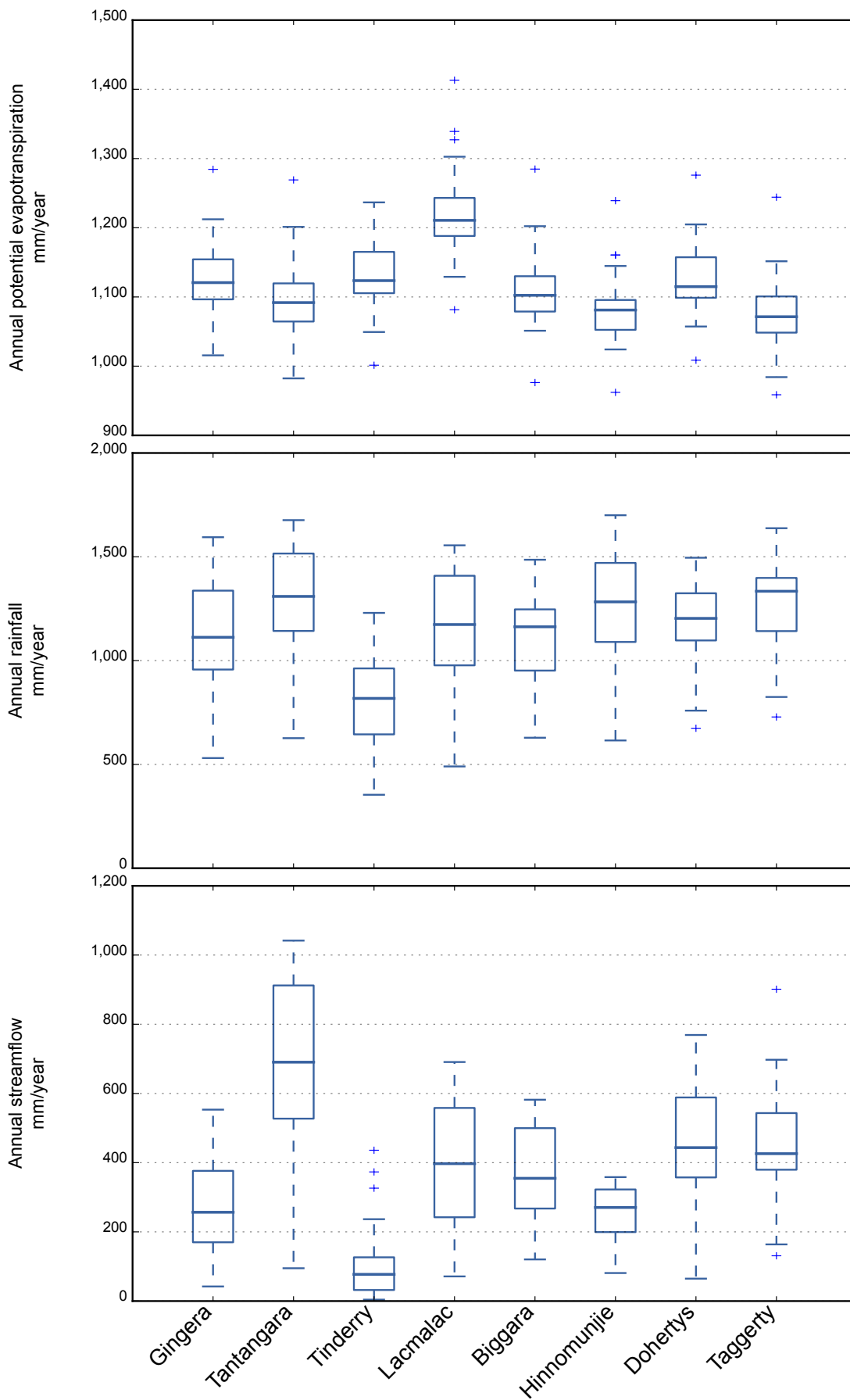
Figure 5: Annual potential evaporation rates, rainfall and runoff of the study catchments
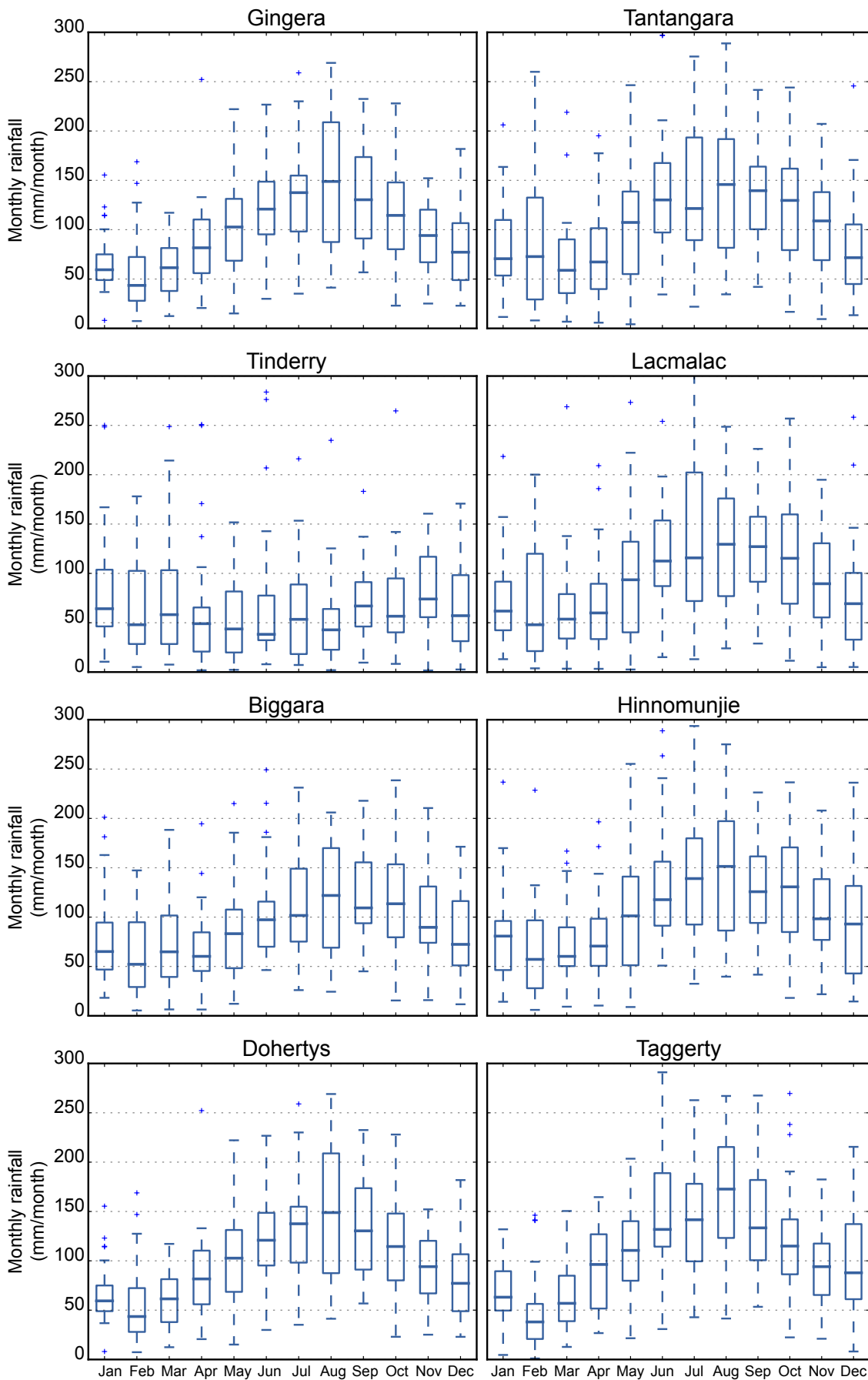
Figure 6: Monthly rainfall of the study catchments
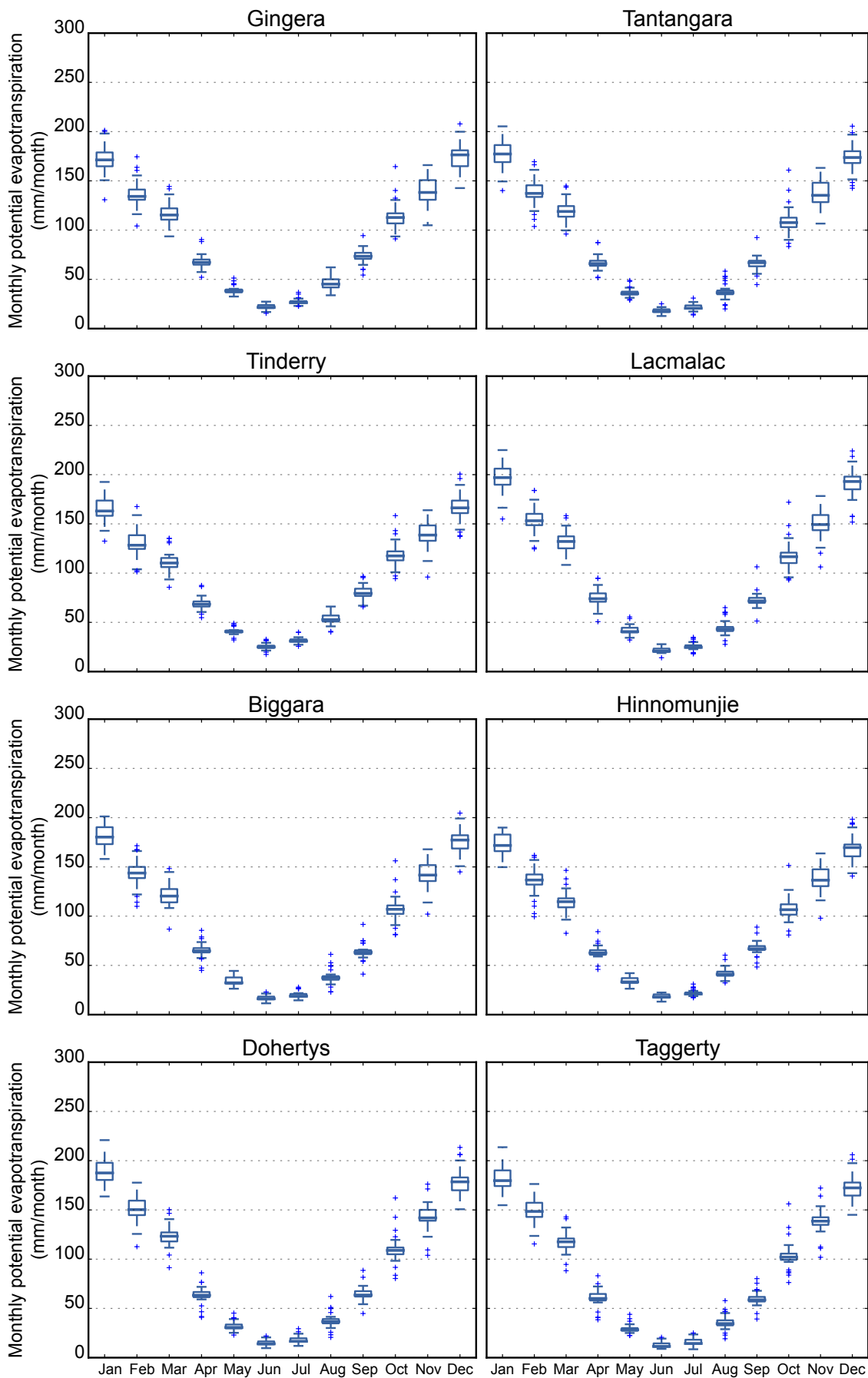
Figure 7: Monthly potential evaporation rates of the study catchments

As noted previously, snowmelt exerts a significant influence particularly in early spring, in determining the magnitude and time of seasonal streamflow for some catchments (Schreider et al. 1997). Long-term average of monthly rainfall and runoff for the Hinnomunjie and Tantangara are shown in Figure 3. Hinnomunjie has an upper area on the border of Mt. Bogong, where elevation is up to 1,940 m above sea level. Snowmelt from the area contributes to the delayed augment of streamflow in early spring. For Tantangara Reservoir, the impact of snowmelt on the catchment water budget is more striking. The long-term average of early spring runoff exceeds that of rainfall derived from the AWAP dataset. It seems obvious that in the Hinnomunjie and Tantangara catchments, a large proportion of the total streamflow in early spring comes from snowmelt in the catchments.

Mean annual rainfall in the target catchments for 1975–2006 vary in the range 799 to 1,273 mm/yr whereas mean annual runoff varies in the range 124 to 683 mm/yr (Table 1). The runoff coefficients of target catchments clearly indicate that most catchments are high yielding catchments. The highest rainfall of 1,273 mm/yr and the highest runoff coefficient of 0.54 correspond to the Tantangara catchment. It is possible that the catchment rainfall may have been somewhat underestimated due mainly to gridding procedures in the AWAP rainfall over steep alpine terrains. Tuteja et al. (2007) found similar inconsistencies in the steep mountainous catchments of the nearby Snowy River. Among the target catchments, Tinderry has the lowest rainfall of 799 mm/yr and also the lowest runoff coefficient of about 0.16.

3.4   Downscaled POAMA rainfall

POAMA (Predictive Ocean Atmosphere Model for Australia: poama.bom.gov.au; Alves et al. 2003) was developed by the Bureau as a state-of-the-art seasonal to interannual forecast system based on a coupled ocean/atmosphere model and ocean/atmosphere/land observation assimilation systems. POAMA demonstrated skills for predicting the variations in Australian rainfall associated with tropical sea surface temperature at lead times to at least one season. The first version (POAMA-1) was developed jointly between the Bureau, the former division of CSIRO Marine Research and the Managing Climate Variability (MCV) program and became operational in October 2002. The main focus for POAMA-1 was the prediction of SST anomalies associated with the El Niño/Southern Oscillation.

A newer version POAMA1.5 was implemented in the Bureau in June 2007, with real-time predictions produced soon after. POAMA-1.5 uses the same coupled model as in POAMA-1 (with some enhancements) and contains a new atmospheric/land initialisation system, developed as part of the SEACI (South Eastern Australian Climate Initiative) project. POAMA-1 and 1.5 are both T47 spectral resolution which is approximately 2.5° x 2.5° or 270 km x 270 km (see Figure 8 for grid resolution of the POAMA and observed AWAP climatology). The ability of POAMA 1.5 to predict large scale drivers of Australian climate and Australian rainfall (directly) was analysed in several studies, mostly as part of the SEACI project (e.g. Hendon & Alves 2009). POAMA-1.5 demonstrated reasonable skill in large scale regional climate variables; however, the predictions of rainfall tend to be too emphatic, i.e. more confident than reality, suggesting that the ensemble spread is not large enough.

For the rainfall downscaling, we used POAMA 1.5 hindcast dataset (forecast data in simulation mode), which provides nine-month forecasts. Forecast datasets are available at daily time steps on the first day of the month out to the next nine months. Analogue downscaling method (Charles et al. 2010; Timbal, Li & Fernandez 2009; Timbal & McAvaney 2001) was applied to derive catchment scale rainfall from regional scale POAMA outputs. The skill of the SDMs was evaluated by comparing reconstructed and observed series using a range of metrics: the first two moments of the series, and the ability to reproduce day-to-day variability, interannual variability, and long-term trends.
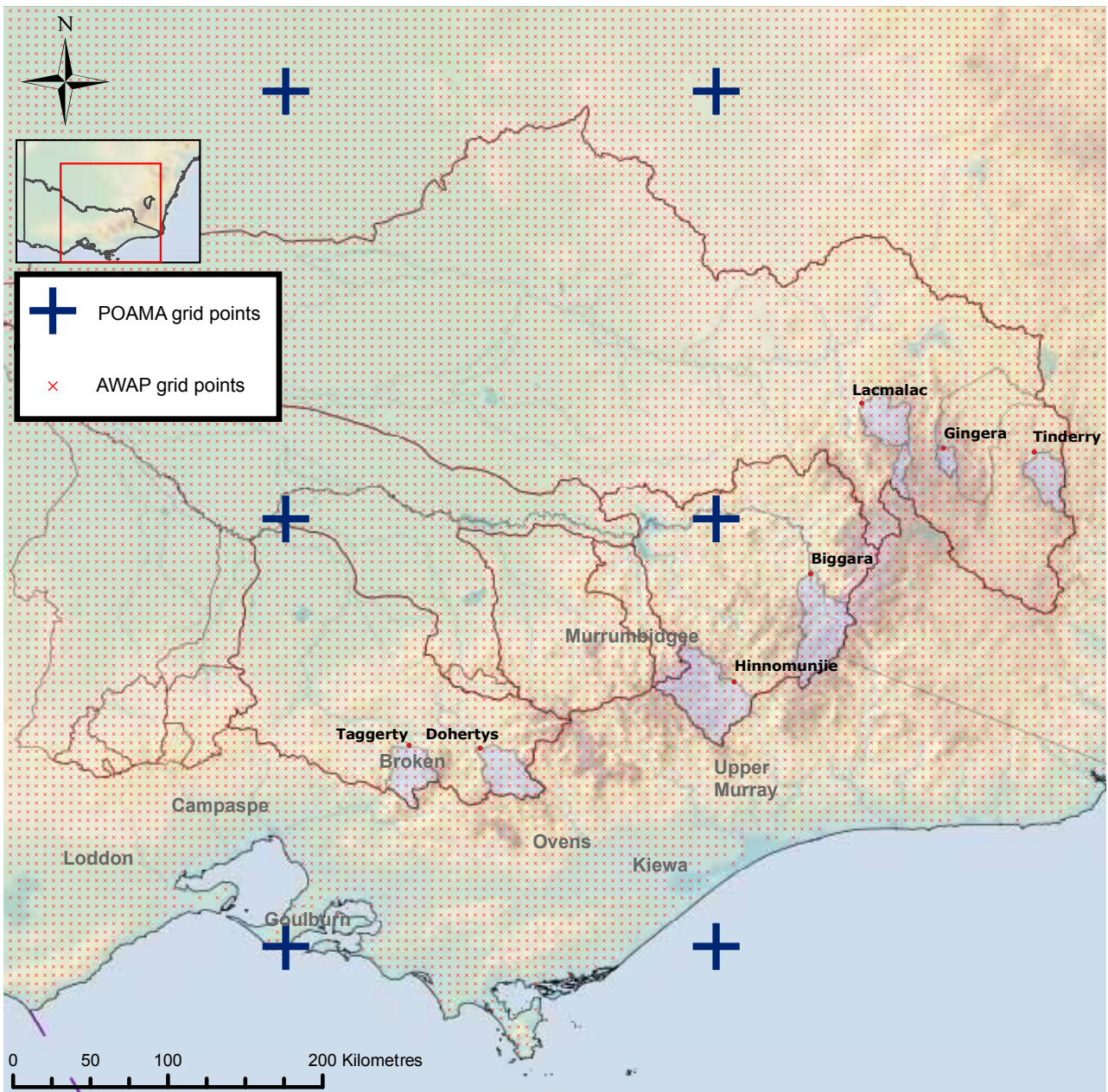
Figure 8: POAMA grid points and AWAP grid points

The predictors used for downscaling are defined as follows:

- Mean Sea Level Pressure (MSLP in hPa)

- surface minimum and maximum temperature (T_{min} and T_{max} in ºC)

- total rainfall (PRCP in mm)

- specific humidity (Q in g/Kg)

- relative humidity (R in %)

- temperature(T in ºC)

- zonal and meridional wind components (U and V in m.s–1).

Timbal, Li and Fernandez (2009) divided the Australian continent into ten regions: Tasmania (TAS), Southwest of Western Australia (SWA), Nullarbor Plain (NUL), the Southwest of Eastern Australia (SEA), the Southern part of the Murray–Darling Basin (SMD), the Southeast Coast (SEC), the Mid-East Coast (MEC), Queensland (QLD), the Northern Monsoon Region (NMR) and the Northwest of Western Australia (NWA). Optimal combinations of predictors for each season and for each of the ten regions were identified in Timbal, Li and Fernandez (2009).

Typically, the direct rainfall output from a GCM (Global Climate Model) is too emphatic and seriously biased. Therefore, it is important to adjust the GCM climatology at regional scale using downscaling for reliable results. Shao and Li (2010; 2011a) assessed the current practice in bias correction of GCM outputs and developed a new bias correction method and applied it to the analogue method of Timbal, Li and Fernandez (2009). The performance was evaluated by testing the downscaling results using an analogue method, with and without bias correction for POAMA. The bias correction on predictors was slightly modified for better accuracy and distribution matching (Shao and Li 2011a). Monthly downscaled POAMA rainfall against observation data for the Gingera and Biggara catchments are compared in Figure 9 and Figure 10.

The relationship between POAMA rainfall and AWAP observed rainfall turned out to be weak even for the monthly scale. Particularly, rainfall during wet months was often underestimated, which considerably reduced the variation range of POAMA rainfall compared to that of observed data. The serious underestimation of rainfall during wet seasons became one major source for bias in forecast streamflow outcomes from the rainfall-runoff models (e.g. see Figure 10).
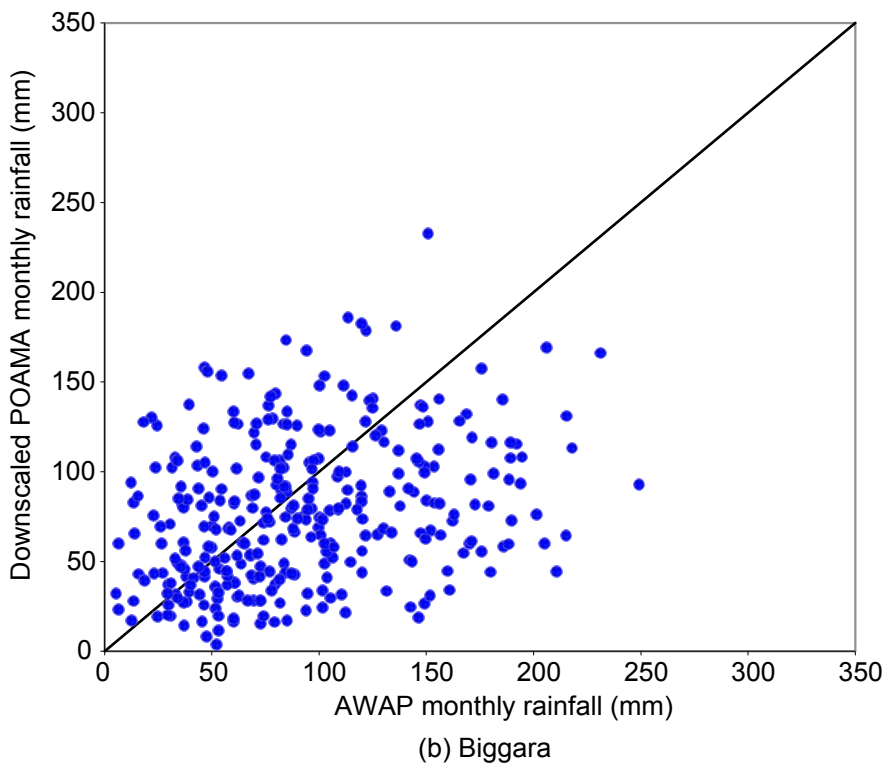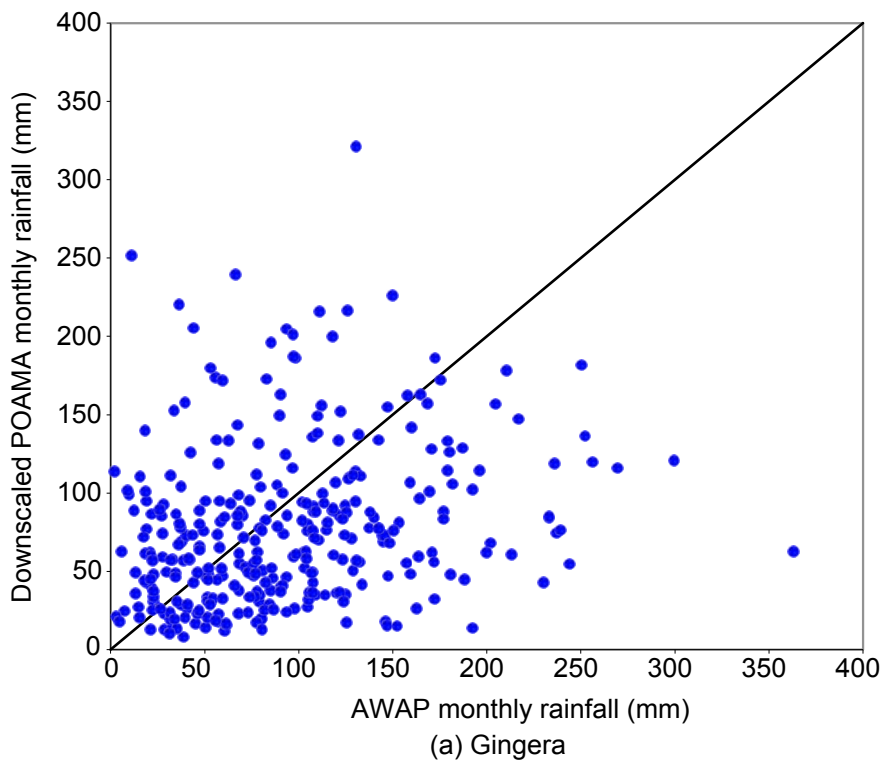
Figure 9: Monthly AWAP rainfall data and POAMA rainfall forecast for Gingera and Biggara
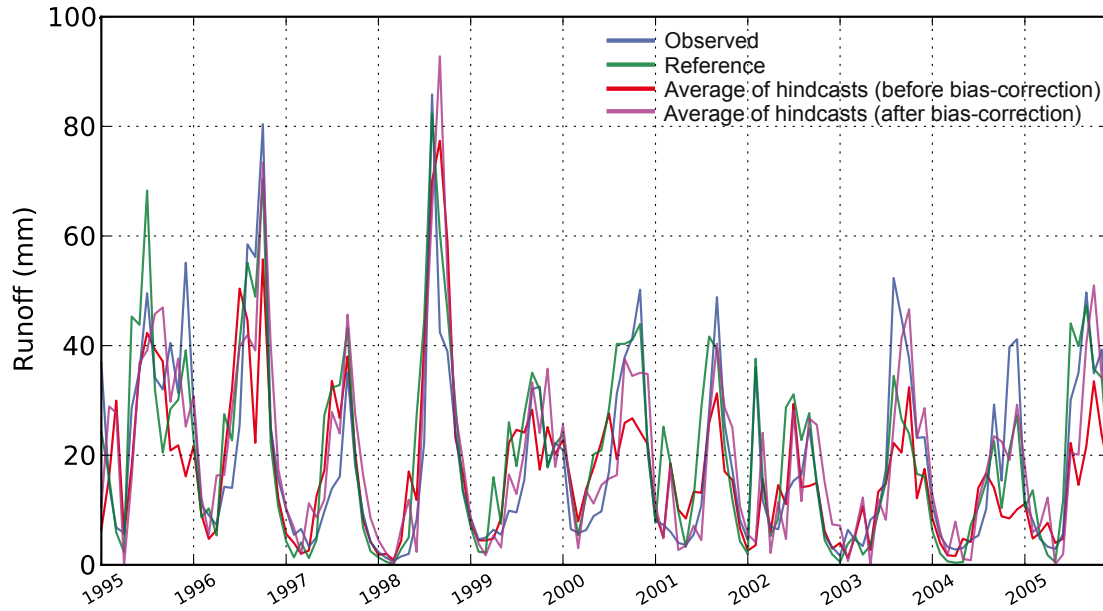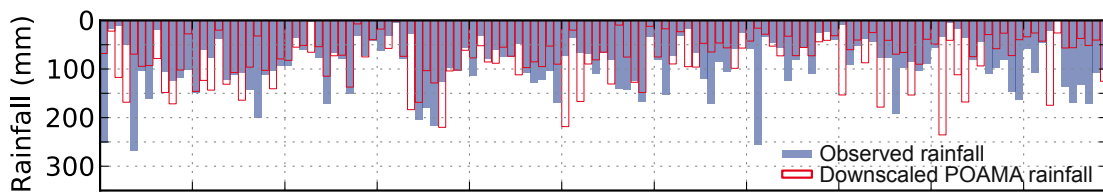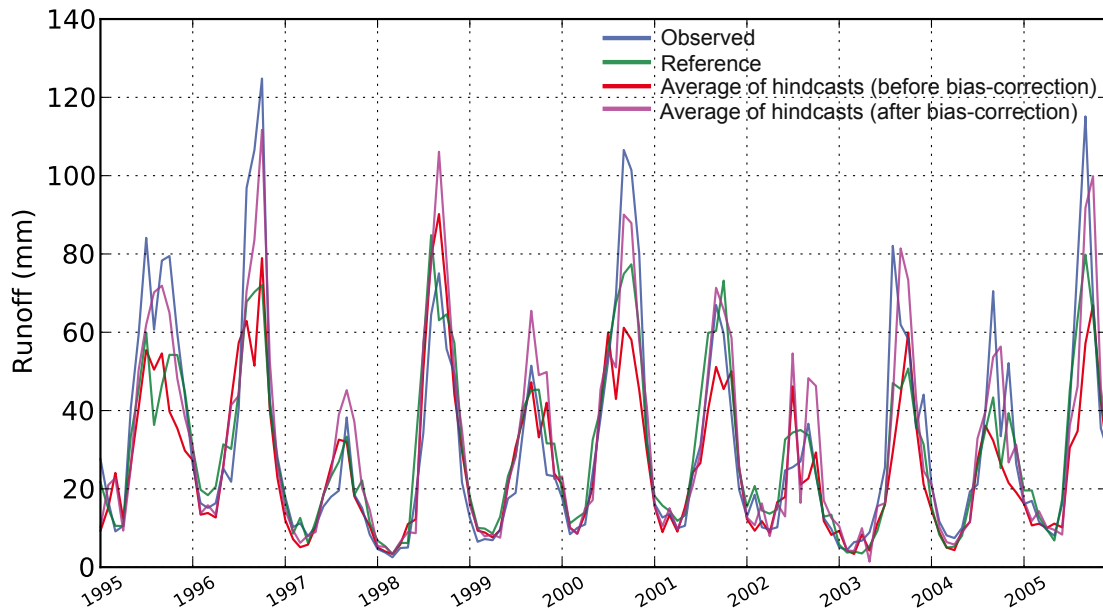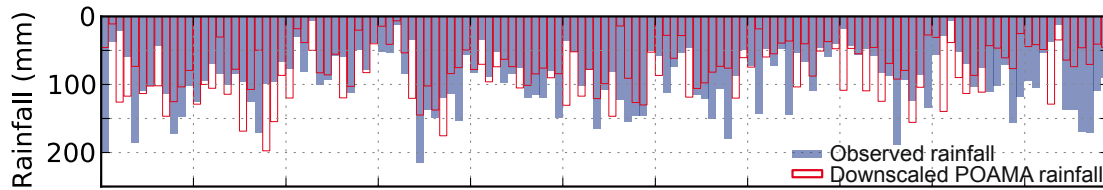
Figure 10: Hyetographs of observed monthly rainfall and downscaled POAMA rainfall and hydrographs of observed and hindcast monthly streamflow for Gingera and Biggara. The hindcast streamflows are monthly ones updated on the first day of each month

4. Methods

4.1   Rainfall-runoff models

The history of hydrologic modelling ranges from the Rational Method of Mulvany (1850) to lumped rainfall-runoff models developed since the 1960s [e.g. Stanford Watershed Model (Crawford & Linsley 1966); Sacramento (Burnash 1985); SMAR (Kachroo & Liang 1992); SIMHYD (Chiew, Zhou & McMahon 2003)]. An integrated, process-based hydrologic modelling approach is, in general, appropriate to investigate the intricate, nonlinear relationships between the land surface, the unsaturated zone, the saturated zone and the river discharge under changing non-stationary conditions, though the large number of parameters complicates using such an approach. In more recent times, distributed/semi-distributed physically-meaningful models have been developed [e.g. TOPMODEL (Beven & Kirkby 1979), DHSVM (Wigmosta, Vail & Lettenmaier 1994); TOPOG (Vertessy & Elsenbeer 1999); TOPKAPI (Liu & Todini 2002); CLASS (Tuteja et al. 2004); REW (Reggiani & Schellekens 2006); Grid-to-Grid (Moore et al. 2006)].

We believe the single biggest unknown in this study is our ability to forecast one-month and three-month ahead rainfall and that the use of more sophisticated semi-distributed or distributed hydrologic models would further complicate the study. Therefore, we decided to use the following well known and established lumped rainfall-runoff models in this evaluation: Sacramento, SIMHYD and SMAR (Podger 2004: www.toolkit.net.au/rrl). They are simple but have stood the test of time since the first conceptual rainfall-runoff model was developed in the 1960s (Chiew et al. 2008; Crawford & Linsley 1966). These models have been applied for both rural and urban water supply catchments. They have been widely used in Australia by governmental agencies and river basin authorities in water resources assessments, catchment water balance studies, water sustainable yields projects and water forecasting operations. All three models use the daily rainfall and pan or potential evapotranspiration data as inputs to simulate daily streamflows.

4.1.1   SIMHYD

SIMHYD (Chiew, Peel & Western 2002) is a conceptual rainfall-runoff model that simulates daily streamflow from daily rainfall and areal potential evapotranspiration data. SIMHYD is a simplified version of the HYDROLOG model that was developed in 1972 (Porter 1972; Porter & McMahon 1975) and the more recent MODHYDROLOG model (Chiew & McMahon 1994). SIMHYD has seven parameters which are required to be calibrated (Figure 11).

The model estimates daily streamflow from three sources: infiltration excess runoff, interflow (and saturation excess runoff) and baseflow. Daily rainfall first fills the interception store, part of which is lost through evaporation. The excess rainfall is then subjected to an infiltration function that determines the infiltration capacity. The excess rainfall that exceeds the infiltration capacity becomes infiltration excess runoff.

Moisture that infiltrates is subjected to a soil moisture function that diverts the water to the stream as interflow, to groundwater store as recharge or to remain in the soil moisture store. Interflow is first estimated as a linear function of the soil wetness. The equation used to simulate interflow therefore attempts to mimic both the interflow and saturation excess runoff processes using the soil wetness to reflect parts of the catchment that are saturated from which saturation excess runoff can occur. Groundwater recharge is then estimated, also as a linear function of the soil wetness. The remaining moisture flows into the soil moisture store.

Evaporation from the soil moisture is estimated as a linear function of the soil wetness, but cannot exceed the atmospheric controlled rate of areal evapotranspiration. The soil moisture store has a finite capacity, part of which percolates into the groundwater store. Baseflow from the groundwater store is simulated as a linear recession from the store.
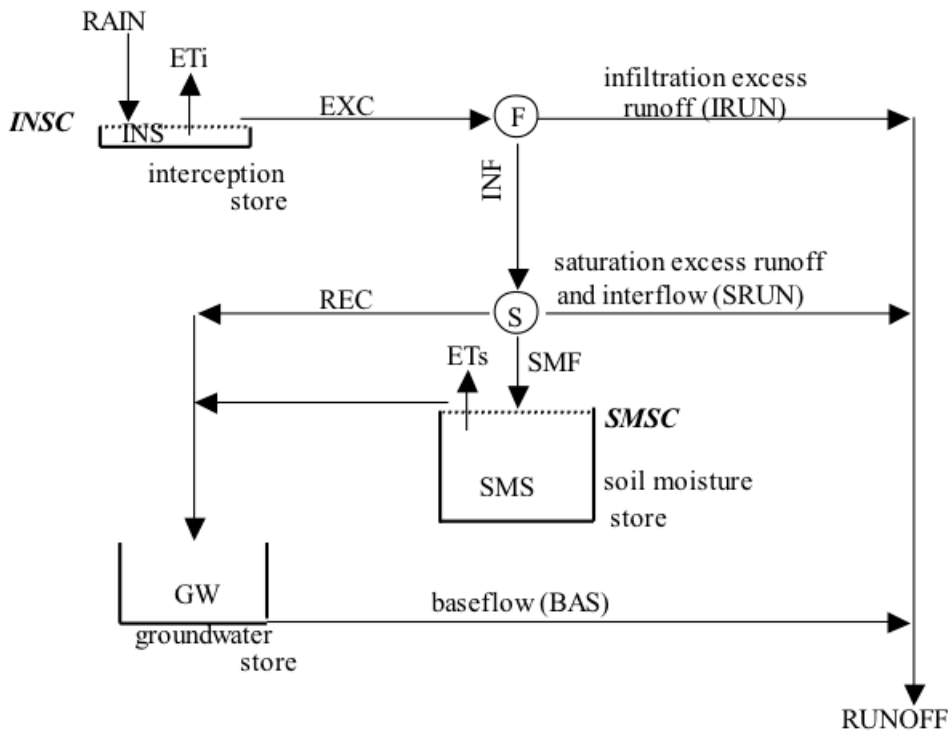
PET = areal potential evapotranspiration (input data)

EXC = max{(RAIN + INS - INSC) , 0.0}

ETi = lesser of (INS, PET)

INF = lesser of {*COEFF* exp(-*SQ* x SMS/*SMSC*), EXC}

IRUN = EXC - INF

SRUN = *SUB* x SMS/*SMSC* x INF

REC = *CRAK* x SMS/*SMSC* x (INF - SRUN)

SMF = INF - SRUN - REC

POT = PET - ETi

ETs = lesser of {10 x SMS/*SMSC* , POT}

BAS = *K* x GW

Model Parameters

INSC   interception store capacity (mm)

COEFF   maximum infiltration loss (mm)

SQ   infiltration loss exponent

SMSC   soil moisture store capacity (mm)

SUB   constant of proportionality in interflow equation

CRAK   constant of proportionality in groundwater recharge eqn.

K   baseflow linear recession parameter

Figure 11: Schematic diagram of SIMHYD model structure (source: Chiew, Peel & Western 2002)

4.1.2 Sacramento

The Sacramento model is a conceptual rainfall-runoff model used to generate daily streamflow from rainfall and evaporation data (Burnash 1985; Burnash et al. 1973: Figure 12).

The model uses soil moisture accounting to simulate the water balance within the catchment. Soil moisture storage is increased by rainfall and reduced by evaporation and by outflow from the storage. The size and relative wetness of the storage then determines the amount of rainfall absorbed, actual evapotranspiration loss, and the amount of water moving vertically or laterally out of the store. The surface runoff is transformed from excess rainfall through an empirical unit hydrograph. This runoff is then added with lateral water movements from the soil moisture stores to estimate streamflow. The Sacramento model uses a total of 22 parameters to simulate water balance.

Of these: five parameters define size of the soil moisture stores (Lzfpm, Lzfsm, Lztwm, Uzfwm and Uztwm), three parameters calculate the rate of lateral outflows (Lzpk, Lzsk and Uzk), three parameters calculate the percolation water from the upper to the lower soil moisture stores (Pfree, Rexp and Zperc), two parameters calculate direct runoff (Adimp and Pctim), four parameters calculate losses in the system (sarva, side, Ssout and Rserv) and five parameters (UH1, UH2, UH3, UH4, UH4 and UH5) simulate streamflow routing through channels.

The Sacramento model represents the moisture distribution within hypothetical zones of a soil column. The model attempts to maintain percolation characteristics to simulate streamflow contributions from a basin. The components of the Sacramento model are tension water, free water, surface flow, lateral drainage, evapotranspiration (ET) and vertical drainage (percolation).
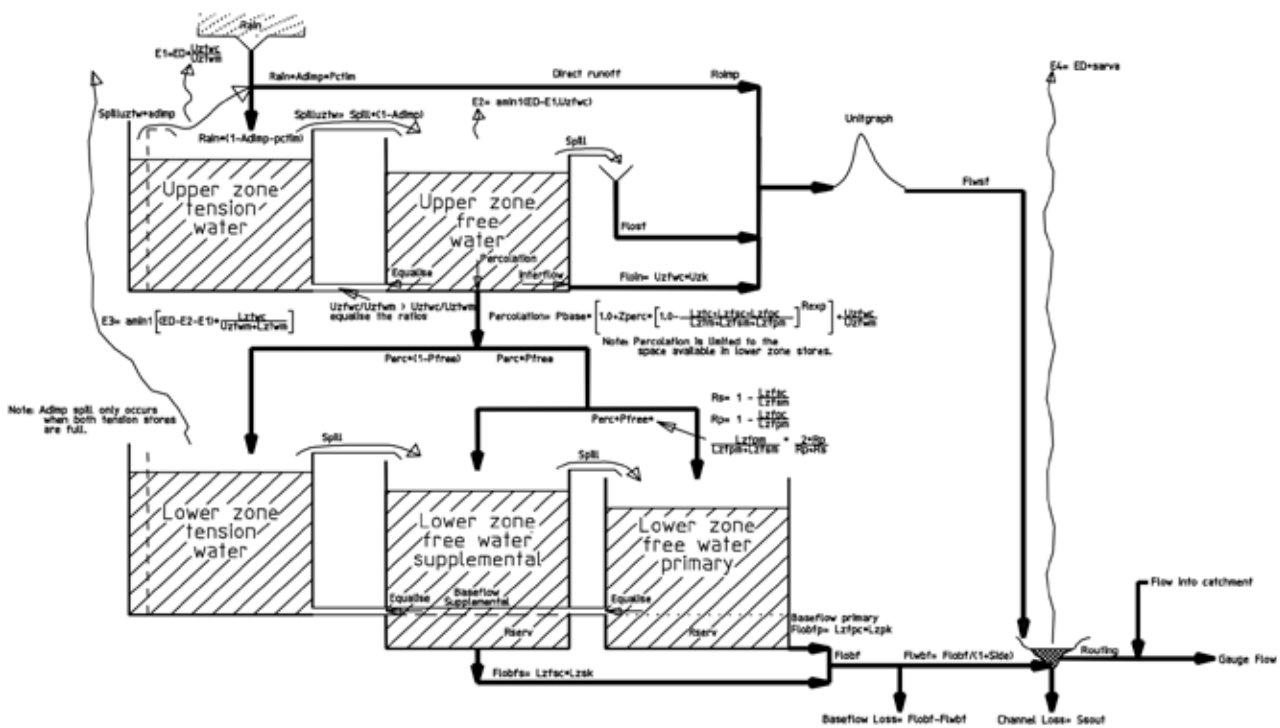
Figure: 12 Schematic diagram of Sacramento model structure (source: Podger 2004)

4.1.3 SMAR

The SMAR soil moisture and accounting model is a lumped conceptual rainfall-runoff water balance model with soil moisture as a central theme (Kachroo 1992; O'Connell, Nash & Farrell 1970; Tuteja & Cunnane 1999: Figure 13).

The model provides daily estimates of surface runoff, groundwater discharge, evapotranspiration and leakage from the soil profile for the catchment as a whole. The surface runoff component comprises overland flow, saturation excess runoff and saturated throughflow from perched groundwater conditions with a quick response time.

The SMAR model consists of two components in sequence, namely, a water balance component and a routing component. The model utilises time series of rainfall and pan evaporation data to simulate streamflow at the catchment outlet. The model is calibrated against observed daily streamflow.

The water balance component divides the soil column into horizontal layers, which contain a prescribed amount of water (usually 25 mm) at their field capacities. Evaporation from soil layers is treated in a way that reduces the soil moisture storage in an exponential manner from a given potential evapotranspiration demand. The routing component transforms the surface runoff generated from the water balance component to the catchment outlet by a gamma function model form (Nash 1960), a parametric solution of the differential routing equation in a single input, single output system. The generated groundwater runoff is routed through a single linear reservoir and provides the groundwater contribution to the stream at the catchment outlet. The SMAR model contains six water balance parameters (C, Z, H, Y, T and G) and three routing parameters (Kg, n and nK).
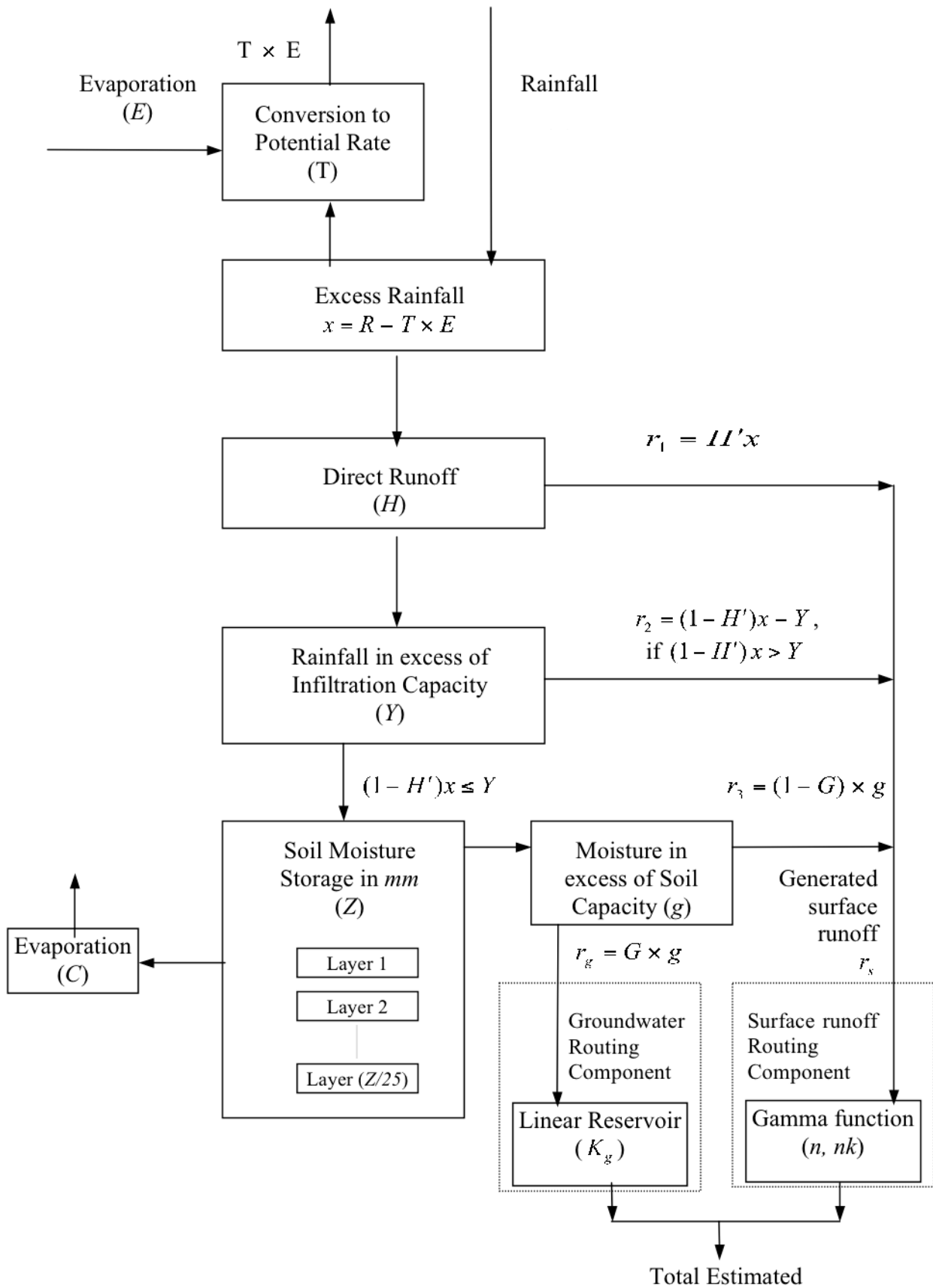
Figure 13: Schematic diagram of SMAR model structure (source: Tuteja & Cunnane 1999)

4.2   Modelling approach

Hydrologic modelling in retrospective mode was performed to satisfy three primary practical requirements: model accuracy (acceptable prediction skills), model consistency (level of accuracy persists through different samples of data) and model versatility (accurate and consistent predictions when subject to diverse applications involving model evaluation criteria not directly based on the objective function used to calibrate the model). A split sample test was adopted for model calibration and validation in retrospective mode. The calibrated and validated hydrological models were then evaluated in forecast simulation mode by using rainfall forcing derived either from downscaled POAMA or, alternatively, from historical rainfall ensembles, which are past rainfall records in same seasons.

4.2.1   Model calibration

The first step for the evaluation was to obtain calibrated parameter values for the three rainfall-runoff models. The models were calibrated using the following methodology:

- the hydrologic models were calibrated for 1976–1996 with 1975 as a warm-up period and validated for 1998–2008 with 1997 as a warm-up period

- as initial seeds, 100 parameter sets were randomly selected from a uniform distribution within the default bound of each parameter

- from the 100 randomly sampled initial parameter sets, optimal parameter values were searched using the Rosenbrock method (Rosenbrock 1960).

Among various objective functions available to evaluate model performance (Krause, Boyle & Bäse 2005; Podger 2004), we chose the objective function specified in equation 1. This objective function is a combination of the Mean Squared Error (MSE) and the error in volumetric ratio between the difference in total volumes of the observed and estimated discharge hydrographs to that of the observed hydrograph (i.e. the index of volumetric fit). The equation turned out to give balanced weights to both high flow and low flow by combining bias penalty with mean square errors.

$$\text{Objective function} = \frac{\sum_{i=1}^{n}\left(Q_{obs,i} - Q_{sim,i}\right)^2}{n} \times \left[1 + \frac{\left|\overline{Q_{obs}} - \overline{Q_{sim}}\right|}{\overline{Q_{obs}}}\right] \quad (1)$$

where:

Q_{obs} = daily observed discharge (mm)

Q_{sim} = daily simulated discharge (mm)

$\overline{Q_{obs}}$ = average daily observed discharge (mm)

$\overline{Q_{sim}}$ = average daily simulated discharge (mm)

n = total length of data in days

To assess the effects of monthly or seasonal variation of the climate on streamflow, two different calibration schemes were tested:

1. General parameterisation scheme to find one optimal parameter set for the entire period.

2. Conditional parameterisation scheme to find 12 optimal parameter sets, each of which corresponds to individual months (Luo et al. 2011 *under review*; Wang, Zheng & Luo 2010; Wang et al. 2011a; Wang et al. 2011b). For instance, one optimal parameter set for January streamflow simulation, one for February and so on. In this case, the value of objective function was optimised for streamflow outcome for each month. Note that each of the 12 model representations was allowed to run continuously in parallel even though each individual model was optimised for a specific target month wherein data for other months was ignored.

Five parameter sets selected by the general parameterisation scheme for Gingera and Biggara catchments are shown in Table 2, Table 3, Table 4, Table 5, Table 6 and Table 7, for three rainfall-runoff models respectively.

Table 2: Top five calibrated parameter sets of SIMHYD for Gingera

Parameter name	Parameter range		1	2	3	4	5
	Min	Max					
baseParameter, K (-)	0	1	0.108	0.111	0.110	0.112	0.112
imperviousThreshold (mm)	0	5	0.676	1.127	5.000	4.838	4.921
maximumInfiltrationLoss, COEFF (-)	0	400	117.536	335.972	309.236	329.982	271.149
infiltrationLossExponent, SQ (-)	0	10	0.253	1.312	0.127	1.195	0.561
interflowCoefficient, SUB (-)	0	1	0.000	0.000	0.000	0.000	0.000
perviousFraction (-)	0	1	0.993	0.996	0.994	0.996	0.997
interceptionStoreCapacity, INSC (mm)	0	5	0.007	0.000	0.000	0.000	0.000
rechargeCoefficient, CRAK (-)	0	1	0.525	0.533	0.536	0.536	0.537
soilMoistureStoreCapacity, SMSC (mm)	1	500	450.464	450.396	451.079	448.256	447.528

Table 3: Top five calibrated parameter sets of SIMHYD for Biggara

Parameter name	Parameter range		1	2	3	4	5
	Min	Max					
baseParameter, K (-)	0	1	0.062	0.071	0.066	0.070	0.071
imperviousThreshold (mm)	0	5	0.000	0.869	4.268	4.495	2.143
maximumInfiltrationLoss, COEFF (-)	0	400	134.101	144.469	95.744	98.633	164.574
infiltrationLossExponent, SQ (-)	0	10	1.090	1.127	0.597	0.619	1.384
interflowCoefficient, SUB (-)	0	1	0.068	0.034	0.079	0.044	0.054
perviousFraction (-)	0	1	0.996	0.984	1.000	0.985	0.992
interceptionStoreCapacity, INSC (mm)	0	5	0.651	0.210	0.567	0.311	0.109
rechargeCoefficient, CRAK (-)	0	1	0.842	0.735	0.895	0.826	0.762
soilMoistureStoreCapacity, SMSC (mm)	1	500	296.189	292.369	335.981	324.969	318.574

Table 4: Top five calibrated parameter sets of Sacramento for Gingera

Parameter name	Parameter range		1	2	3	4	5
	Min	Max					
Adimp (-)	0.000	1.000	0.000	0.000	0.000	0.000	0.000
Lzfpm (mm)	0.000	50.000	37.696	45.360	49.433	3.691	50.000
Lzfsm (mm)	0.000	50.000	18.897	6.283	49.548	49.948	4.909
Lzpk (1/day)	0.000	1.000	0.034	0.037	0.009	0.404	0.036
Lzsk (1/day)	0.000	1.000	0.016	0.031	0.031	0.028	0.028
Lztwm (mm)	0.000	400.000	393.233	382.738	326.958	365.454	397.519
Pctim (-)	0.000	1.000	0.017	0.012	0.003	0.017	0.013
Pfree (-)	0.000	1.000	0.938	1.000	0.987	0.779	1.000
Rexp (-)	0.000	3.000	0.853	0.914	1.522	0.565	1.020
Rserv (-)	0.000	1.000	0.040	0.003	0.332	0.000	0.000
Sarva (-)	0.000	1.000	0.000	0.000	0.000	0.000	0.000
Side (-)	0.000	1.000	0.035	0.170	0.315	0.171	0.300
Ssout (m3/s/km2)	0.000	1.000	0.000	0.000	0.003	0.000	0.000
UH1 (-)	0.000	1.000	0.900	0.900	0.900	0.900	0.900
UH2 (-)	0.000	1.000	0.100	0.100	0.100	0.100	0.100
UH3 (-)	0.000	1.000	0.000	0.000	0.000	0.000	0.000
UH4 (-)	0.000	1.000	0.000	0.000	0.000	0.000	0.000
UH5 (-)	0.000	1.000	0.000	0.000	0.000	0.000	0.000
Uzfwm (mm)	0.000	80.000	79.976	77.600	78.784	76.346	80.000
Uzk (1/day)	0.000	1.000	0.074	0.071	0.103	0.049	0.061
Uztwm (mm)	0.000	100.000	71.328	47.846	48.415	38.022	39.880
Zperc (-)	0.000	80.000	65.739	46.795	69.818	22.384	30.697

Table 5: Top five calibrated parameter sets of Sacramento for Biggara

Parameter name	Parameter range		1	2	3	4	5
	Min	Max					
Adimp (-)	0.000	1.000	0.016	0.023	0.005	0.049	0.001
Lzfpm (mm)	0.000	50.000	41.517	44.952	26.389	50.000	26.370
Lzfsm (mm)	0.000	50.000	48.674	47.324	31.462	41.587	50.000
Lzpk (1/day)	0.000	1.000	0.004	0.011	0.006	0.013	0.011
Lzsk (1/day)	0.000	1.000	0.035	0.035	0.065	0.035	0.037
Lztwm (mm)	0.000	400.000	303.086	275.270	357.419	249.225	286.366
Pctim (-)	0.000	1.000	0.026	0.026	0.034	0.022	0.042
Pfree (-)	0.000	1.000	1.000	1.000	1.000	1.000	1.000
Rexp (-)	0.000	3.000	1.186	1.559	1.420	2.178	1.824
Rserv (-)	0.000	1.000	0.057	0.000	0.002	0.007	0.000
sarva (-)	0.000	1.000	0.000	0.000	0.000	0.000	0.000
Side (-)	0.000	1.000	0.000	0.030	0.000	0.069	0.000
Ssout (m3/s/km2)	0.000	1.000	0.000	0.000	0.000	0.000	0.018
UH1 (-)	0.000	1.000	0.900	0.900	0.900	0.900	0.900
UH2 (-)	0.000	1.000	0.100	0.100	0.100	0.100	0.100
UH3 (-)	0.000	1.000	0.000	0.000	0.000	0.000	0.000
UH4 (-)	0.000	1.000	0.000	0.000	0.000	0.000	0.000
UH5 (-)	0.000	1.000	0.000	0.000	0.000	0.000	0.000
Uzfwm (mm)	0.000	80.000	63.951	59.106	61.748	78.879	62.197
Uzk (1/day)	0.000	1.000	0.060	0.069	0.052	0.052	0.059
Uztwm (mm)	0.000	100.000	7.522	17.473	6.142	18.613	23.841
Zperc (-)	0.000	80.000	59.913	63.040	79.478	80.000	66.325

Table 6: Top five calibrated parameter sets of SMAR for Gingera

Parameter name	Parameter range		1	2	3	4	5
	Min	Max					
Evaporation Coeff - C (-)	0	1	1	1	1	1	1
Groundwater runoff coeff - G (-)	0	1	1	1	1	1	1
Direct runoff area index - H (-)	0	1	0.22	0.22	0.22	0.22	0.22
Groundwater parameter - Kg (-)	0.01	200	178.06	178.05	178.11	178.02	178
Parameter n of Nash model - n (-)	1	10	1	1	1	1	1
Parameter - nk (-)	1	10	2.9	2.9	2.9	2.9	2.9
Potential evaporation factor - T (-)	1	1	1	1	1	1	1
Soil infiltration capacity - Y (mms/T.S)	100	100	100	100	100	100	100
Total soil moisture capacity - Z (mms)	0	125	125	125	125	125	125

Table 7: Top five calibrated parameter sets of SMAR for Biggara

Parameter name	Parameter range		1	2	3	4	5
	Min	Max					
Evaporation Coeff - C (-)	0	1	0.47	0.47	0.47	0.47	0.47
Groundwater runoff coeff - G (-)	0	1	1	1	1	1	1
Direct runoff area index - H (-)	0	1	0.15	0.15	0.15	0.16	0.15
Groundwater parameter - Kg (-)	0.01	200	69.33	69.24	69.8	69.84	69.55
Parameter n of Nash model - n (-)	1	10	1	1	1	1	1
Parameter - nk (-)	1	10	1.27	1.28	1.3	1.35	1.27
Potential evaporation factor - T (-)	1	1	1	1	1	1	1
Soil infiltration capacity - Y (mms/T.S)	100	100	100	100	100	100	100
Total soil moisture capacity - Z (mms)	0	125	125	125	125	125	125

4.2.2   Forecast

Forecasting in simulation mode or hindcasting is the most important component for evaluating any forecasting approach. It involves retrospective forecasting using past predictand information to re-forecast (or hindcast) known events so as to evaluate a forecasting strategy. By comparing the streamflow forecast in simulation mode with observed data, we can evaluate the likely forecast quality in the actual forecast mode.

Forecasting in simulation mode was performed for 1985–2005 with a warm-up period of 1980–1984 (Figure 14). Although downscaled POAMA rainfall forecasts (hindcasts) are available for 1980–2006, we chose 1985–2005 as the forecast period to select the same record length as that of the historical rainfall ensemble approach. The hydrologic models were calibrated for 1976–1996 with 1975 as a warm-up period and validated for 1998–2008 with 1997 as a warm-up period. Therefore, forecasting is performed

using data for the period already used for hydrologic model calibration. While acknowledging that the overlap between calibration and forecasting period should be avoided, we are somewhat constrained due to common period of forecast data possible between downscaled POAMA 1.5 (1980–2006) and historical rainfall ensemble (1985–2005). Nevertheless, we believe that the presence of a number of wet and extremely dry periods in the calibration and validation periods respectively, resulted in hydrologic model representations that can be confidently used in forecasting. This limitation would be overcome in the next phase of this work prior to operationalisation when POAMA 2.4a–c datasets for 1960–2006 would be used for forecasting in simulation mode. POAMA 2.4a–c implementation is based upon better assimilation of ocean and atmospheric conditions relative to POAMA 1.5 even though resolution of the new model is same. These forecast datasets from POAMA 2.4 were not available during this evaluation project.

As rainfall input, data from the following two sources were used in simulations:

1. downscaled POAMA rainfall ensemble (ten members and their mean)

2. historical rainfall ensemble (Wang et al. 2011), which is the past ten-year observed rainfall data (ten members and their mean).

PET data from AWAP for the respective period were used because the interannual variation and the effects of PET on total streamflow outcome are negligible compared to those from rainfall input. Parameter sets from both parameterisation schemes were used to set up the hydrologic models. In addition, two types of streamflow forecasts with different temporal scale were generated:

• seasonal (three-monthly) streamflow forecasts updated every first day of the month

• monthly streamflow forecasts updated every fortnight i.e. first and 16th day of the month.
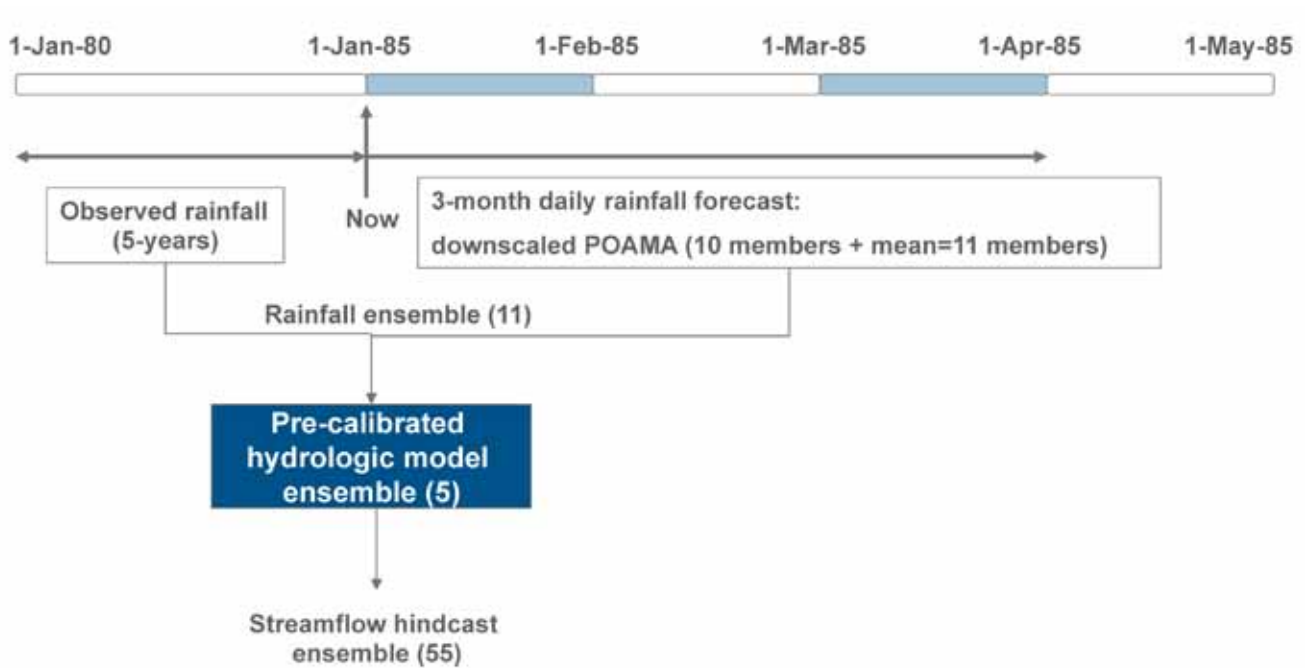
Figure 14: Schematic diagram to show the forecast scheme in simulation mode

The following forecast procedure was used in simulation mode (Figure 14):

- For each forecast update, initialise a rainfall-runoff model with the past five-year daily rainfall and PET data. Through model re-initialisation driven by observed data, the internal storages of the model were re-calculated at each time step to represent the antecedent soil moisture and groundwater conditions properly.

- Use downscaled POAMA rainfall ensemble or historical rainfall ensemble as input to the model and simulate three-month ahead daily streamflow forecast. Historical rainfall ensemble is the collection of the past rainfall records observed in the same seasons.

- Aggregate daily streamflow forecasts to seasonal (three-month) and monthly streamflow forecasts.

For each hydrologic model, with five different parameter sets and 11 different rainfall ensemble members (i.e. ten downscaled POAMA ensembles and ensemble mean as the 11th member), we obtained a total of 55 streamflow forecast ensemble members at each time step.

4.2.3 Streamflow bias correction

Bias correction is a procedure to correct the tendency for the simulated runoff to be systematically larger or smaller than the observed runoff. This bias could be caused by model input errors (observed rainfall and potential evapotranspiration), model structural deficiency and poorly calibrated model parameters, problems with the numerical implementations, runoff measurement errors and, more importantly, inaccurate rainfall forecasts (see section 7 on BATEA for full treatment of predictive uncertainty) (Kavetski, Kuczera & Franks 2006a; Kavetski, Kuczera & Franks 2006b; Kuczera et al. 2010a; Kuczera et al. 2010b).

Regardless of the cause, when a bias is present it will lower the performance in forecast simulation as well as in the real-time forecast modes. From a process perspective, it is desirable to maximise forecast skills before any posterior bias correction is applied directly on streamflow forecasts. However, in most practical hydrologic and weather forecasting problems, this is rarely the case and the need for a bias correction procedure is often there, albeit to a different degree depending on the type of problem.

Assuming that correcting the underlying cause of this bias was intractable, a bias correction process was tested using the Autoregressive Moving Average (ARMA: Box & Jenkins 1976; Hyndman 2011) model. An ARMA model described by equations 2 and 3 is a linear filter model that converts a sequence of uncorrelated (usually normal) random variables ε_t (or the white noise) to a sequence of correlated variables (e.g. streamflow bias time series).

$$\left(1 - \sum_{i=1}^{p} \phi_i B^i \right) y_t = \left(1 + \sum_{j=1}^{q} \theta_j B^j \right) \varepsilon_t \quad (2)$$

$$\Phi(B) y_t = \Theta(B) \varepsilon_t \quad (3)$$

where:

$i = 1, 2, \ldots p = $ order of the autoregressive AR component

$j = 1, 2, \ldots q = $ order of the moving average MA component

y_t = streamflow bias time series (observed flow minus median runoff from the 55 member ensemble of the forecast runoff

ε_t = standard normal variate N (0, σ_ε^2)

σ_ε^2 = residual variance i.e. remaining variance after bias correction

$B = $ backward shift operator,

$$\Phi(B) = 1 - \sum_{i=1}^{p} \phi_i B^i \,,$$

$$\Theta(B) = 1 + \sum_{j=1}^{q} \theta_j B^j$$

A common difficulty in implementation of the ARMA model is that the order of the autoregressive and moving average components is usually considered subjective and somewhat difficult to apply. There have been several attempts to automate ARMA modelling over the last 25 years. Hannan and Rissanen (1982) proposed a method to identify the order of an ARMA model for a stationary time series by fitting a long autoregressive model to the data, and then the likelihood of potential models was computed via a series of standard regressions. Gómez (1998) extended the Hannan–Rissanen identification method to include multiplicative seasonal ARMA model identification. Gómez and Maravall (1997) implemented this automatic identification procedure in the software TRAMO and SEATS. Other examples of automatic methods include the use of filtering methods and certain heuristic rules (Goodrich 2000; Liu 1989; Makridakis & Hibon 2000; Ord & Lowe 1996; Reilly 2000).

We used the automatic time series forecasting method of Hyndman and Khandakar (2008). The ARMA model is calibrated according to the Akaike Information Criterion AIC (p,q) which is dependent on the maximum likelihood function L and order of the autoregressive and moving average components of the ARMA model:

$$AIC = -2\log(L) + 2(p+q) \quad (4)$$

The likelihood function is defined by the mean squared error (MSE) or $\sum (\varepsilon_t)^2 / n$ from equation 2. There are several constraints on the fitted models to avoid the problems with convergence or near unit-roots:

- The values of p and q are not allowed to exceed specified upper bounds (with default values of 5 in each case).

- A model is rejected if it is 'close' to non-invertible (when roots of $\Theta(B) = 0$ lie outside the unit circle) or non-stationary (when roots of $\Phi(B) = 0$ lie outside the unit circle).

- If either of the characteristic equation $\Theta(B) = 0$ or $\Phi(B) = 0$ have a root smaller than 1.001 in absolute value, the model is rejected.

- If there are any errors arising in the non-linear optimisation routine used for estimation, the model is rejected. The rationale used here is that any model that is difficult to fit is probably not a good model for the data.

With the constraints, the algorithm guarantees to return a valid model because the model space is finite and at least one of the starting models will be accepted; i.e. the model without any autoregressive and moving average parameters.

To remove the autocorrelation structure remaining in the residuals of the forecast streamflow, the following ARMA-based bias correction was applied on streamflow outcome (see Appendix A for the details).

1. Calculate the median runoff from the 55 member ensemble of a single forecast runoff outcome.

2. Calculate the residual between the median and observed runoff.

3. Standardise the residual time series for each forecast update period using the respective mean and standard deviation.

4. Determine and calibrate the best order of an ARMA model using the standardised residual time series.

5. Apply the calibrated ARMA model to simulate new standardised residuals.

6. De-standardise the simulated residuals and estimate bias correction required in the (biased) median forecast runoff.

7. Apply the simulated residuals to the median runoff and each of the runoff ensemble members from step 1. Note that through this step, bias correction in the biased median forecast runoff is applied to the entire biased forecast distribution.

8. Store this ensemble of bias corrected runoff.

The methodology was implemented as a DMS Utility Tool (Laugesen 2010) using the R programming language (R Development Core Team 2011) and the Forecast package (robjhyndman.com/software/forecast). The bias corrected procedure was implemented using biased forecast distributions for 1985–2005 (21 years) in the following manner:

- *Split sampling scheme 1*: Calibrate an ARMA model with data of 1985–1998 (14 years) and correct the bias of streamflow outcome of 2000–2005 (six years) to validate the ARMA model. Means and standard deviations for 1985–1998 were used for 2000–2005 in steps 3 and 6 above.

- *Split sampling scheme 2*: Calibrate an ARMA model with data of 1992–2005 (14 years) and correct the bias of streamflow outcome of 1985–1991 (seven years) to validate the ARMA model. Means and standard deviations for 1992–2005 were used for 1985–1991 in steps 3 and 6 above.

- *Combined calibration scheme*: If outcomes from the two split sample tests above lead to comparable model performance in the respective validation period, then re-calibrate the ARMA model for the period 1985–2005 (21 years) to obtain bias corrected forecast distribution for evaluation using the skill scores discussed in section 4.3.

Note that this last step was done to include enough samples for statistically meaningful estimates of forecast verification. Since the split samples include periods of distinctly different hydrometeorological conditions, the results are expected to provide conservative estimates about the likely improvement we can have with the bias correction. Because of splitting, we could not have enough samples for skill scores and reliability calculation. Therefore, we report only NSE values of the split sample test results. As the final results of forecast verification, we evaluate skill scores using four indices and assess reliability of the bias corrected probabilistic forecasts using the predictive QQ plots, which are explained in the following section.

4.3   Forecast verification

4.3.1   Skill scores

Forecast skill is defined as 'the relative accuracy of a set of forecasts, with respect to some set of standard control or reference forecasts.' (Wilks 1995, p. 237) In many cases, the distribution or its average of past predictand data were used as reference forecasts. In this study, we chose past observed streamflow data as reference forecasts.

Forecast skill is often quantified by a skill score. The generic form of skill score is:

$$Skill\ Score = \frac{Score_{fcast} - Score_{ref}}{Score_{perf} - Score_{ref}} \times 100\ (\%) \quad (5)$$

where:

$Score_{fcast}$ is the score of streamflow forecasts

$Score_{ref}$ is the score of reference forecasts

$Score_{perf}$ is the score of perfect forecasts

The lower bound of the skill score varies depending on a selected score, but the upper bound of the skill score is 100 (%) for any score. It is noteworthy that the actual value of a skill score can be different even for the same data, depending on reference forecasts selected for the calculation.

This study used three skill scores: RMSE, RMSEP and CRPS. For all three scores, the score of perfect forecasts $Score_{perf}$ is zero. These skill scores were derived using the same code that is currently being used in the operational statistical seasonal forecasting system (www.bom.gov.au/water/ssf). When calculating these skill scores for a time step, runoff value observed at the time step was not included to avoid any artificial increase of skills scores.

The skill score of RMSE (Root Mean-Squared Error) is the distance between the median of streamflow forecasts and corresponding observations. It is calculated by:

$$RMSE_{fcast} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(median(\hat{Q}_{i,k}) - Q_{i,k}\right)^2} \ , \quad (6)$$

$$RMSE_{ref} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(median(\tilde{Q}_{i,k}) - Q_{i,k}\right)^2} \ , \quad (7)$$

where:

n is the number of years of the forecast period in simulation mode (e.g. equals 21 years for 1985–2005)

k is the update time step (e.g. 12 monthly updates of three-monthly flow in a year)

$\hat{Q}_{i,k}$ is the streamflow forecast in year i and update time step k

$\tilde{Q}_{i,k}$ is the reference forecasts (observed streamflow)

$Q_{i,k}$ is the observation at the kth update

Note that to calculate an RMSE skill score for the kth update, n different cumulative distribution functions were developed, but the streamflow value observed in a given year i was left out when estimating $median(\tilde{Q}_{i,k})$ for the kth update in year i.

The RMSE calculates errors directly in the measurement space. The square power in the equations makes the error term in RMSE sensitive to a few large errors relative to many small ones, and therefore it can potentially lead to conservative forecasts.

RMSEP (Root Mean-Squared Error of Prediction) is the expected value of the distance between the median of streamflow forecasts and corresponding observations in probability space (Wang & Robertson 2011).

It is calculated by:

$$RMSEP_{fcast} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[\tilde{F}\left(median(\hat{Q}_{i,k})\right) - \tilde{F}(Q_{i,k})\right]^2}, \quad (8)$$

$$RMSEP_{ref} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[\tilde{F}\left(median(\tilde{Q}_{i,k})\right) - \tilde{F}(Q_{i,k})\right]^2} \quad (9)$$

where:

$\tilde{F}$ is the cumulative distribution function of reference forecasts

Note that the terms on the right hand side of equations 8 and 9 are in probability domain while the analogous terms in equations 6 and 7 are in the original flow domain. Therefore, the influence of large errors in the original flow domain has a smaller impact on RMSEP relative to RMSE. In the case of RMSEP, more frequent events have more influence on the errors, which reduces the impact of outliers in measurement space. However, for periods with a narrow range of streamflow volumes, such as low flow regimes, a slight error in measurement space can be amplified to an extremely large error in probability space. In such a case, RMSEP can respond sensitively to even marginal errors in streamflow volumes.

CRPS (Continuous Rank Probability Score) is the area between the forecast distribution and a step function H of the observation. It is calculated by:

$$CRPS_{fcast} = \frac{1}{n}\sum_{i=1}^{n}\left[\int_{-\infty}^{\infty}\left[\hat{F}\left(\hat{Q}_{i,k}\right) - H\left(\hat{Q}_{i,k} - Q_{i,k}\right)\right]^2 d\hat{Q}_k\right] , (10)$$

$$CRPS_{ref} = \frac{1}{n}\sum_{i=1}^{n}\left[\int_{-\infty}^{\infty}\left[\tilde{F}\left(\tilde{Q}_{i,k}\right) - H\left(\tilde{Q}_{i,k} - Q_{i,k}\right)\right]^2 d\tilde{Q}_k\right] , (11)$$

where:

$\tilde{F}$ and $\hat{F}$ are the cumulative distribution functions of the reference (observed) and forecast streamflow respectively

$$H\left(Q - Q_{i,k}\right) = \begin{cases} 0, & Q < Q_{i,k} \\ 1, & Q \geq Q_{i,k} \end{cases} \quad (11)$$

Note that for a deterministic forecast, the CRPS is reduced to MAE (Mean Absolute Error) and, like RMSE, it is also sensitive to large errors.

Since each skill score emphasises different aspect of forecast accuracy, it is desirable to inspect all the skill score results when comparing modelling results. In addition to the three skill scores, we also included Nash–Sutcliffe Efficiency (NSE) (Nash & Sutcliffe 1970) because the efficiency measure has been widely adopted in rainfall-runoff modelling. We calculated NSE for the median of streamflow forecasts.

4.3.2 Reliability

Assessment of the reliability of streamflow forecast requires a diagnostic approach that compares a time-varying forecast distribution at all times to a time series of observations. This is a more stringent test than alternative validation methods based on deterministic approaches currently used in hydrology, which simply compare time series of observations and 'optimal' simulations or the standard goodness-of-fit assessments like the Nash–Sutcliffe statistic that cannot check if the forecast distribution is consistent with the observed data.

The forecast distribution is conditioned on the assumptions made during the inference. Consequently, unsupported assumptions may lead to inadequate forecast distributions. Therefore, the estimated forecast distributions must be scrutinised ('validated') to assess reliability of the forecasts (Thyer et al. 2009). Reliability is the agreement between the distribution of forecasts and the distribution of observations. The reliability diagram was used as a visual summarised measure of forecast reliability. In this study, we use the predictive QQ plot, also known as the Probability Integral Transform (PIT) plot, adapted from the verification tools used for probabilistic forecasts of hydrological and meteorological variables (Dawid 1984; De Gooijer & Zerom 2000; Gneiting, Balabdaoui & Raftery 2007; Laio & Tamea 2007; Thyer et al. 2009). The diagram plots the frequency distribution of observations against that of forecasts. The proximity of the plotted curve to the diagonal line in the plot indicates the level of reliability of the forecasts.

For probabilistic forecasts, instead of forecast or observed values, the following PIT values were plotted:

$$\pi_k = \hat{F}\left(Q_k\right) \quad (12)$$

where:

π_k is the PIT for the observation Q_k (Figure 15)

Using the PIT estimates, the predictive QQ plot is constructed as follows (Thyer et al. 2009): Let $\hat{F}$ be the CDF (cumulative distribution function) of the forecast distribution in a given year i and update time step k and $Q_{i,k}$ is the corresponding observed runoff. If the hypotheses in the calibration framework are consistent with the data, the observed value $Q_{i,k}$ should be consistent with the distribution $\hat{F}(Q_{i,k})$.

Hence, under the assumption that the observation $Q_{i,k}$ is a realisation from the forecast distribution, the PIT value $\hat{F}(Q_{i,k})$ is a realisation from a uniform distribution on [0,1]. The predictive QQ plot compares the empirical CDF of the sample of PIT values ($\hat{F}(Q_{i,k}), i = 1,2,....n$) with the CDF of a uniform distribution to assess whether the hypotheses are consistent with the observations.

The predictive QQ plot can be interpreted as follows (Figure 16): (1) If all points fall on the 1:1 line, the forecast distribution agrees perfectly with the observations. (2) If an observed PIT value is 1.0 or 0.0, the corresponding observed data lie outside the forecast range, implying that the predictive uncertainty is significantly underestimated. (3) If the observed PIT values cluster around the midrange (i.e. a low slope around theoretical quartile 0.4–0.6), the predictive uncertainty is overestimated. (4) If the observed PIT values cluster around the tails (i.e. a high slope around theoretical quartile 0.4–0.6), the predictive uncertainty is underestimated. (5) If the observed PIT values at the theoretical median are higher/lower than the theoretical quartiles, the modelled predictions systematically under predict/over predict the observed data.

The predictive QQ plot provides a simple and informative summary of the performance of probabilistic forecasts. It does not involve any additional assumption beyond those used during the calibration and it is a direct test of reliability of the forecasts.
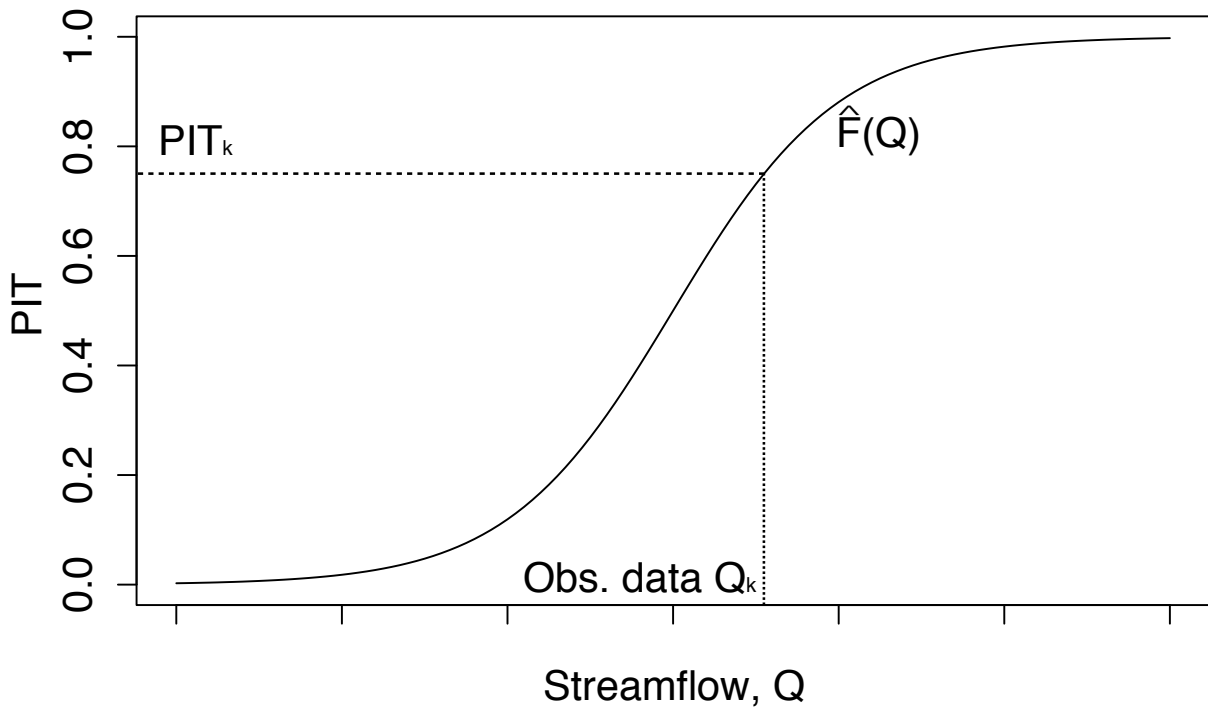
Figure 15: An example of PIT calculation for observed streamflow for the forecast update step k
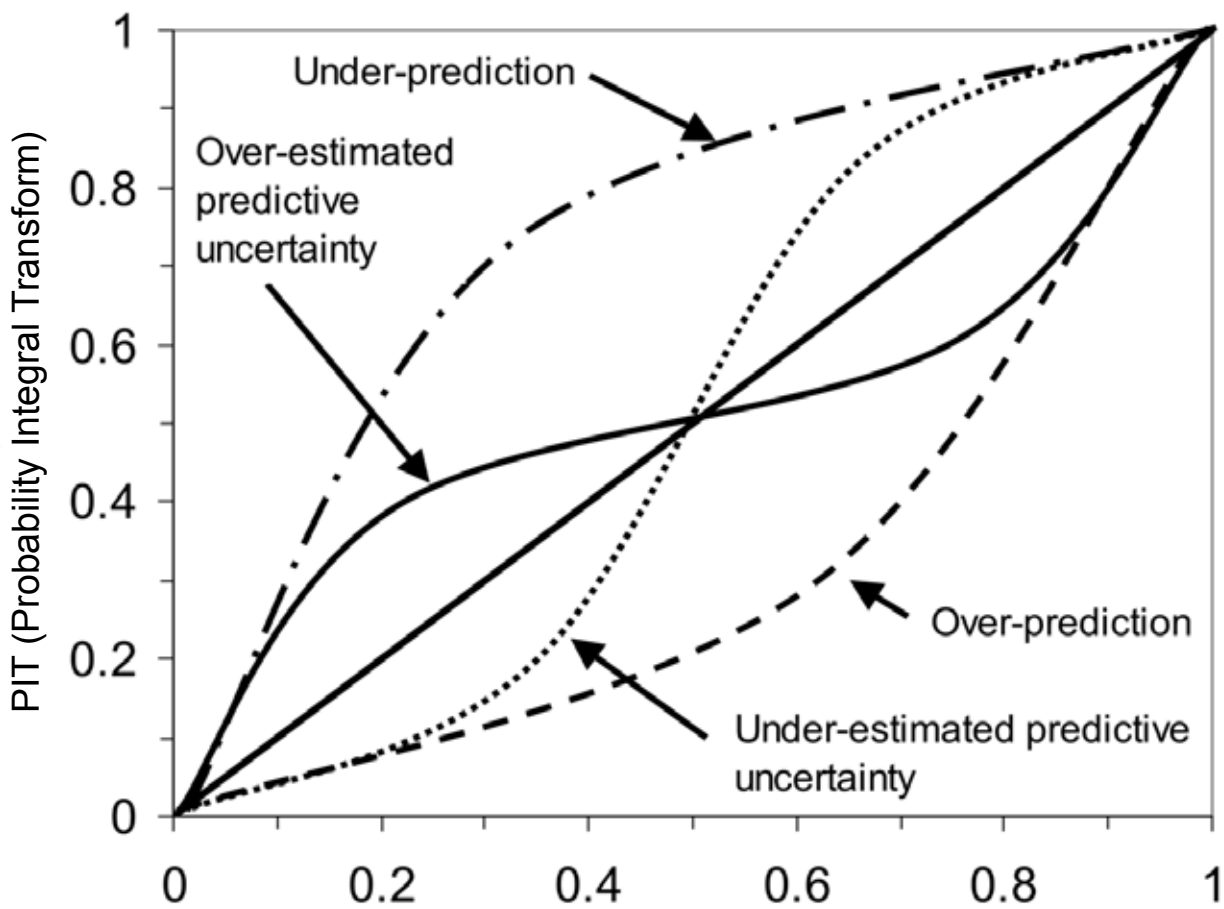
Figure 16: An example of predictive QQ plots to show the PITs against a uniform distribution of observed data (source: Thyer et al. 2009)

5. Modelling system

The Dynamic Modelling System (DMS) was developed as a system for generating streamflow predictions using conceptual rainfall-runoff models. This system allows users to perform model calibration, validation and forecasting; using a workflow based approach (Figure 17). The DMS makes it possible for these functions to be performed in ensemble mode over many combinations of models, input forcings, optimisers, objective functions and accumulation methods for the catchments described section 3 to generate all the outputs (and substantially more) discussed in section 6.

The DMS is an implementation of the recommended solution plan outlined in the Pilot Seasonal Dynamic Modelling System Specifications and Proposed Solution document (Laugesen, Shin & Tuteja 2009). This captured the requirements to meet the deliverables of the Experimental Dynamic Seasonal Streamflow Forecasting project (DM). Structurally, it is a system of interacting software components, which accept input, process it and generate output. This output is then used as input for another component or treated as final output and interpreted by end-users. The structure diagram, illustrated in Figure 17, shows the three main components; Dynamic Modelling Controller (DMC), a large number of utility tools (UT) and a Workflow Manager (WFM).

Each component of the DMS was designed to leverage technologies which suited its purpose, constraints and requirements. These were:

- DMC – C# .NET, eWater TIME (www.toolkit.net.au)

- UT – Python, R and various libraries

- WFM – Python, JSON.

The DMC uses an object-oriented approach to implement the model-view-controller, delegate, bridge and adapter patterns. Each UT uses a structured programming approach with significant commonality between most of the UT. The WFM implements a finite-state machine using object-oriented and structured approaches. All components write log files for debugging purposes, have help documents available at the command-line and are documented in the Dynamic Modelling System Manual (Laugesen 2010).

During the experimental phase, the DMS was primarily operated through a workflow driven process. Hydrologists created high level workflow input files which describe the tasks to perform, where the required input data were located and where the output data should be stored. These workflow input files were then processed by the WFM which ran each task in sequence by launching the requested component through a system-call: either the DMC or a specific UT. Each workflow ran over many combinations of catchment, model, optimiser, objective function, accumulation method, etc. and so were quite complex in some cases. Adopting a workflow-based approach significantly improved the repeatability of complex processes, removed a large amount of human error and this enhanced confidence, accountability and the ability to audit result sets. Figure 18 illustrates a typical workflow driven process.

In addition to driving the DMS through workflows, all individual components could be operated as stand-alone software applications in their own right. Being able to run the DMC or individual UT as stand-alone applications resulted in a significant boost in development parallelisation, component level testing, component re-use and workflow trouble shooting. It came at the cost of slightly lower performance when running in workflow mode, due to the overhead of system context switching and slower data transfer between components, but this was more than made up for by the increase in concurrency and automation. This dual method of operating the DMS was a consequence of adopting a simple architecture relying on common command-line interfaces to all components and a common text file format for transfer and storage of all inputs and outputs. Unfortunately, the use of file-based data persistence resulted in a large and complex directory tree of datasets that made data management difficult in some cases, but this was offset by the benefits of readily accessible results through standard tools and an increased ease of collaboration with our partners from WIRADA, University of Newcastle and the University of Adelaide.

The current implementation of the DMS met the needs for the experimental phase of work for this evaluation project; but it may require further improvement on some components to use DMS as a modelling system for an operational service. In the next phase of this work, these operationalisation issues will be investigated in light of the new or different requirements for an operational system and the lessons learnt from this experimental system.
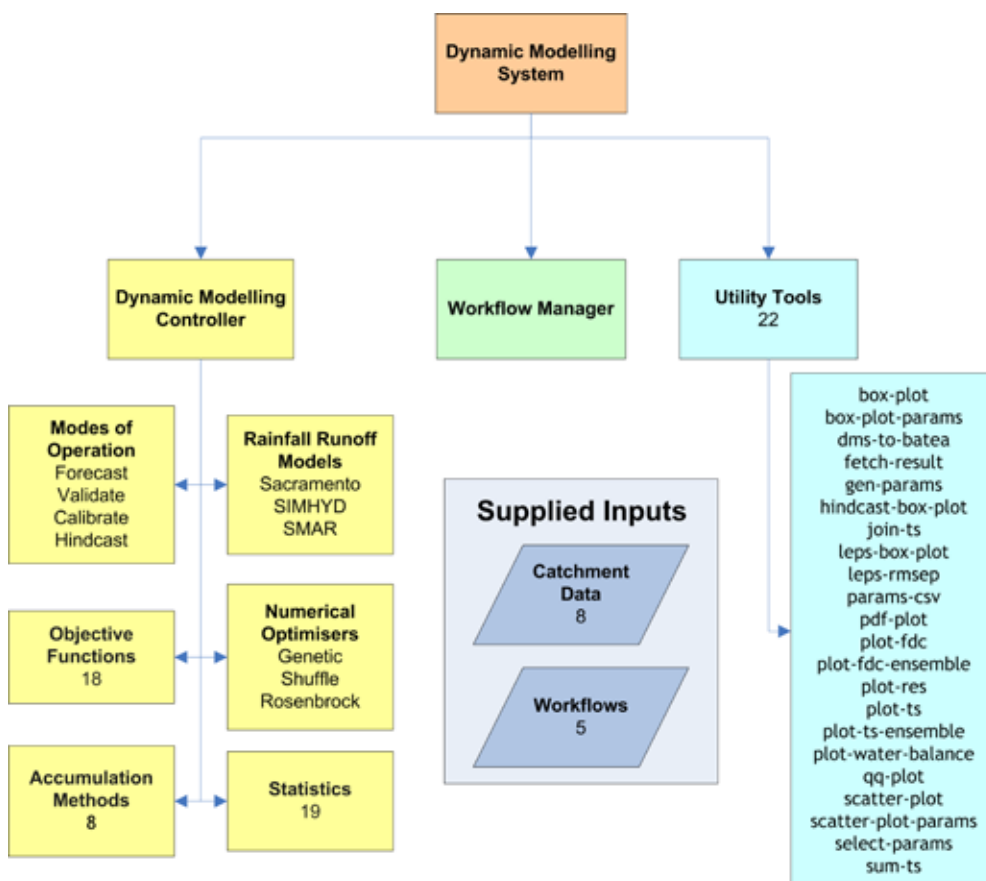
Figure 17: Overview structure of dynamic modelling system (DMS)
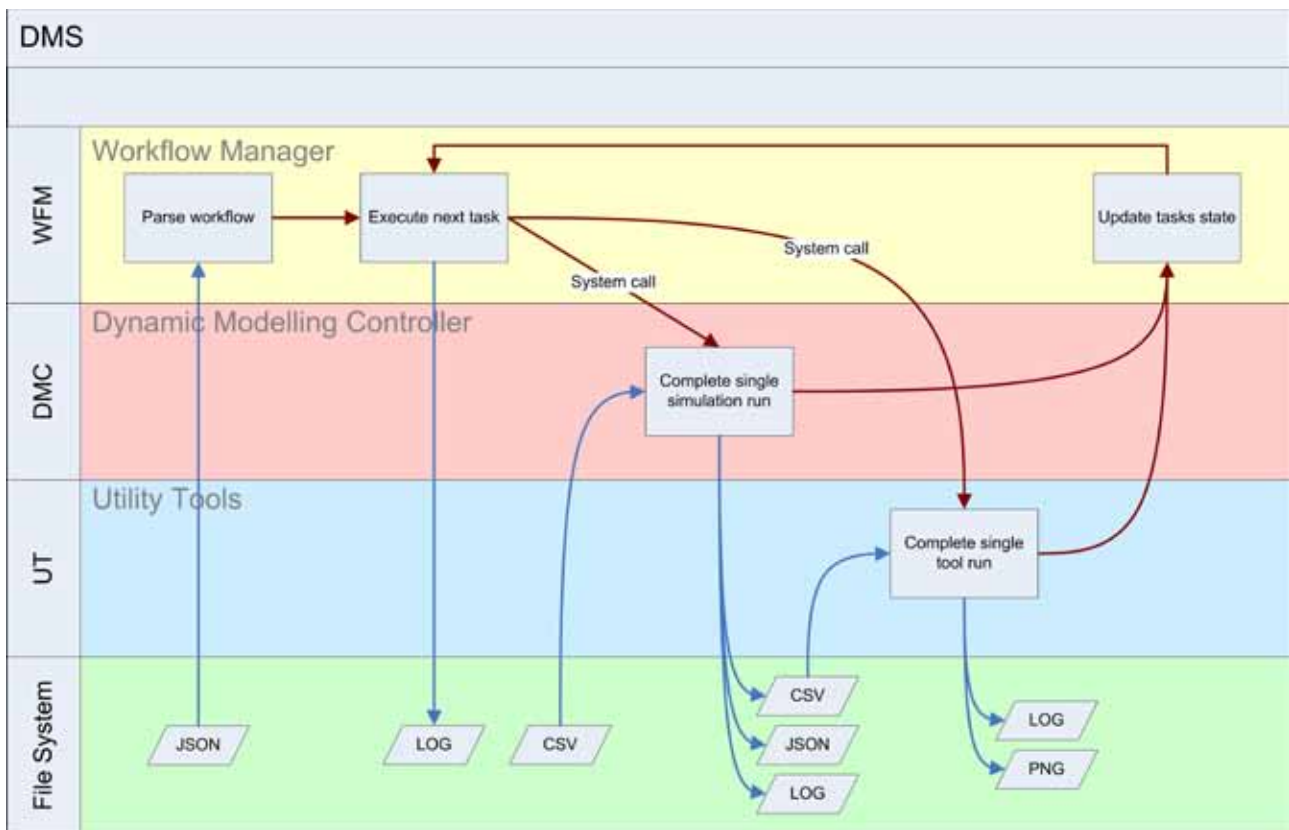
Figure 18: An example of workflow in DMS

6. Results, discussion and evaluation

Modelling methods discussed in section 4 were implemented on the target experimental catchments under the following simulation conditions to evaluate streamflow simulation and forecasting capabilities:

- historical modelling products and retrospective water balance from three rainfall-runoff models in ensemble mode

- streamflow forecasts using hydrologic models derived from general parameterisation scheme versus conditional parameterisation calibration scheme

- streamflow forecasts derived from historical rainfall ensembles versus those from downscaled POAMA rainfall ensembles

- posterior ARMA-based bias correction of the streamflow forecasts

- comparison of streamflow forecasts derived from the dynamic modelling approach versus those from the statistical BJP approach

- monthly streamflow forecast versus seasonal (three-monthly) streamflow forecasts derived from the dynamic approach.

6.1 Comparison of rainfall-runoff models

Performance of the hydrologic model ensembles using observed rainfall data and the suitability of each model for rainfall-runoff prediction is summarised in Table 8 and Figure 19. Monthly streamflow simulated from SIMHYD and Sacramento had NSE averages higher than 75% both for calibration and validation periods in most catchments. The range of NSE values from the SIMHYD model for the calibration and validation periods across all catchments vary in the range 0.52–0.92 and 0.61–0.91 respectively. The respective values from Sacramento vary in the range 0.71–0.95 and 0.66–0.94. Compared to the SIMHYD and Sacramento models, SMAR showed inferior performance for all the catchments (Figure 19).

Table 8: Nash–Sutcliffe efficiency of historical modelling driven by observed rainfall for target catchments

Comparison of NSE (monthly data basis) for eight target catchments				
	Nash–Sutcliffe Efficiency (Optimized on the basis of mean square error and volume constraint: Cal. 1976–96 & Ver. 1998–2008)			
Catchment	Sacramento		Simhyd	
	Calibration	Verification	Calibration	Verification
Gingera	0.772–0.784	0.708–0.732	0.823–0.826	0.802–0.805
Tantangara	0.772–0.832	0.770–0.824	0.804–0.804	0.811–0.812
Tinderry	0.776–0.801	-2.199 to -0.863	0.754–0.758	-1.429 to -1.190
Hinnomunjie	0.707–0.732	0.655–0.737	0.5235–0.5236	0.6085–0.6088
Biggara	0.830–0.849	0.844–0.850	0.821–0.823	0.835–0.836
Dohertys	0.943–0.947	0.872–0.886	0.943–0.943	0.913–0.914
Taggerty	0.935–0.940	0.913–0.938	0.922–0.924	0.867–0.872
Lacmalac	0.903–0.915	0.880–0.893	0.866–0.868	0.805–0.807
Overall range	0.71–0.95	0.66–0.94	0.52–0.92	0.61–0.91
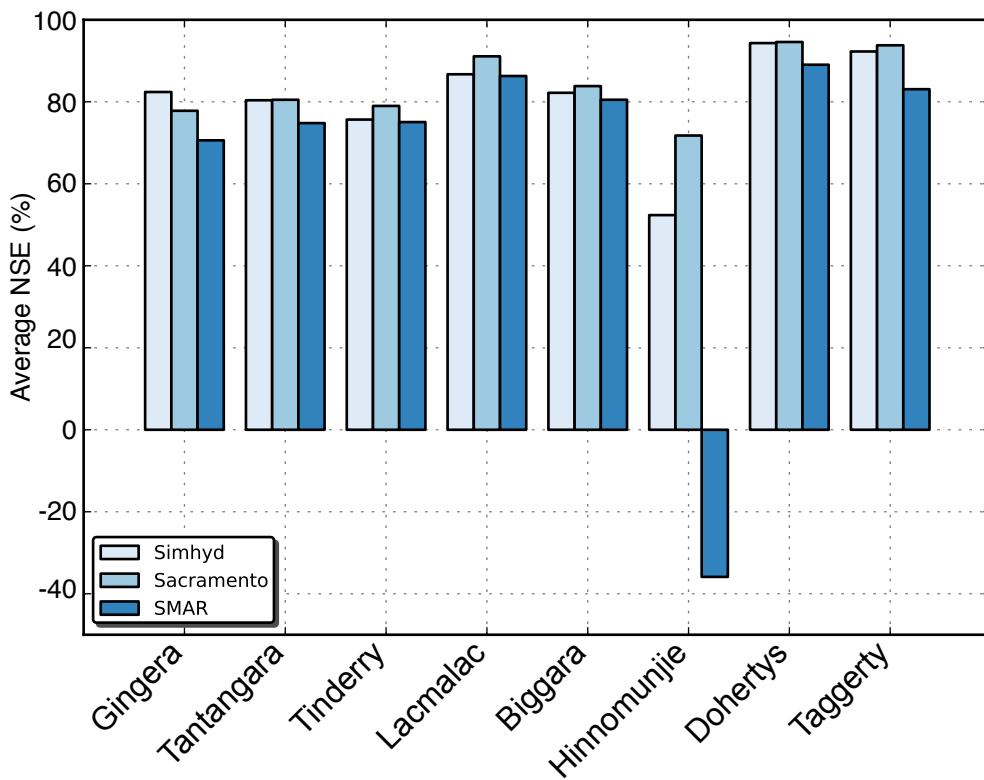
(a) Calibration
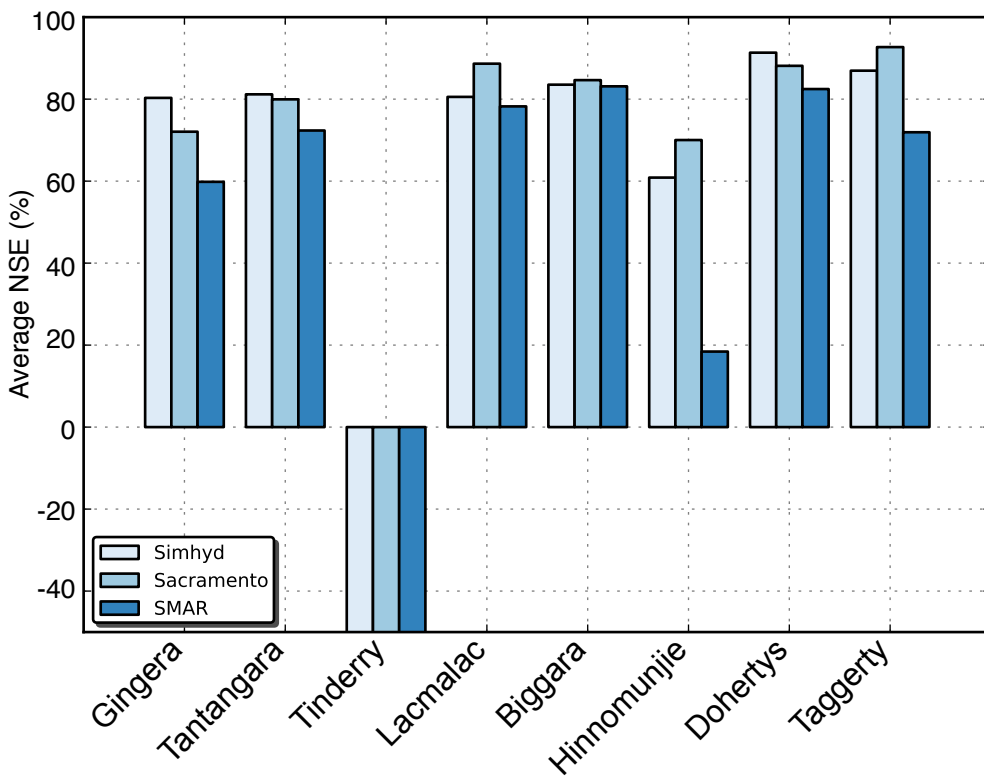
(b) Validation

Figure 19: Averages of Nash–Sutcliffe efficiency of seasonal streamflow outcome from rainfall-runoff models in the calibration period and validation period

For Hinnomunjie, all three models failed to reproduce its monthly streamflow patterns. For Tinderry, none of the models were able to reproduce streamflow patterns observed in the validation period, even though the models were calibrated to simulate streamflow patterns quite accurately during the calibration period.

Figure 20 shows the top 20 monthly streamflow ensemble members of the Biggara catchment. Sacramento produced streamflow ensemble members, each of which had a distinct pattern. However, all 20 streamflow members from SIMHYD had almost the same pattern. Therefore, the use of the Sacramento model in ensemble mode is advantageous whereas the use of the hydrologic model ensembles in the case of SIMHYD is somewhat redundant.

As mentioned before, snowmelt is a significant component in the water budget of the Hinnomunjie and Tantangara catchments. Noting that none of the models includes accounting for snowmelt processes, late winter and early spring runoff in these catchments is often under predicted.

In the case of the Tinderry catchment, all the models generate too much runoff during the dry validation period (1998–2008), because their structures were tuned to produce streamflow patterns during a relatively wetter calibration period (1976–1996). Then, why was the bias so serious only in Tinderry? Note that Tinderry is only one intermittent catchment amongst the target catchments. The catchment has the least mean annual rainfall of 799 mm/yr and the least runoff coefficient of 0.16 amongst the target catchments (Table 1); all other catchments have perennial stream channels with mean annual rainfall in excess of 1,100 mm/yr and runoff coefficients greater than 0.2. For a relatively dry catchment, the representation of nonlinear dynamics in the catchment hydrologic processes, including the interaction between groundwater and stream channel, is critical for determining the streamflow pattern, particularly during a drought. However, it is somewhat difficult to accurately simulate catchment hydrologic processes using models containing only simple approximations of surface and groundwater hydrologic processes, particularly during prolonged dry periods.

Considering the simple structures of models for groundwater flow, it is not surprising that any model could not satisfactorily reproduce streamflow patterns of the dry catchment during droughts. Methodologies required for incorporation of groundwater losses and well level data in the conceptual rainfall-runoff models applied to ephemeral catchments such as those illustrated by Moore and Bell (2002) using the PDM model may need to be considered in the future.

In spite of apparently high NSE values, a closer look at model outcomes reveals a serious limitation of SIMHYD. As an example the high runoff yielding Gingera catchment, which contains steep hill slopes and thin soil cover, is known to produce considerable overland flow and shallow sub-surface flow through the soil (i.e. saturation excess surface runoff). In our simulation, however, SIMHYD generated 98.7% of total streamflow as groundwater flow in the catchment (Figure 21). In other words, SIMHYD reproduced the spiky daily pattern of total streamflow with over-estimated groundwater flux. The quick response of groundwater flux does not conform well to current understanding of the groundwater flow systems in these catchments. Similar model behaviour was also found in other experimental catchments. Further, strong interactions between parameters for the groundwater storage caused the 20 optimal parameter sets to produce almost identical streamflow patterns from the model. Note that a more rigorous evaluation of the SIMHYD model on experimental catchments with BATEA using the Markov Chain Monte Carlo (MCMC) approach confirmed these findings (see section 7).

In summary, historical modelling results highlight the structural deficiency problems of the conceptual rainfall-runoff models, particularly the groundwater component in SIMHYD model, and the implementation error of the SMAR model in the rainfall-runoff model library. In the case of catchments impacted by snowmelt in late winter and early spring, streamflow simulation capability was sub-optimal due to the absence of the snowmelt component in the adopted hydrologic models. Among the three models, Sacramento was selected as the best one that can produce uncertainty bounds of streamflow forecasts with acceptable accuracy for perennial catchments.
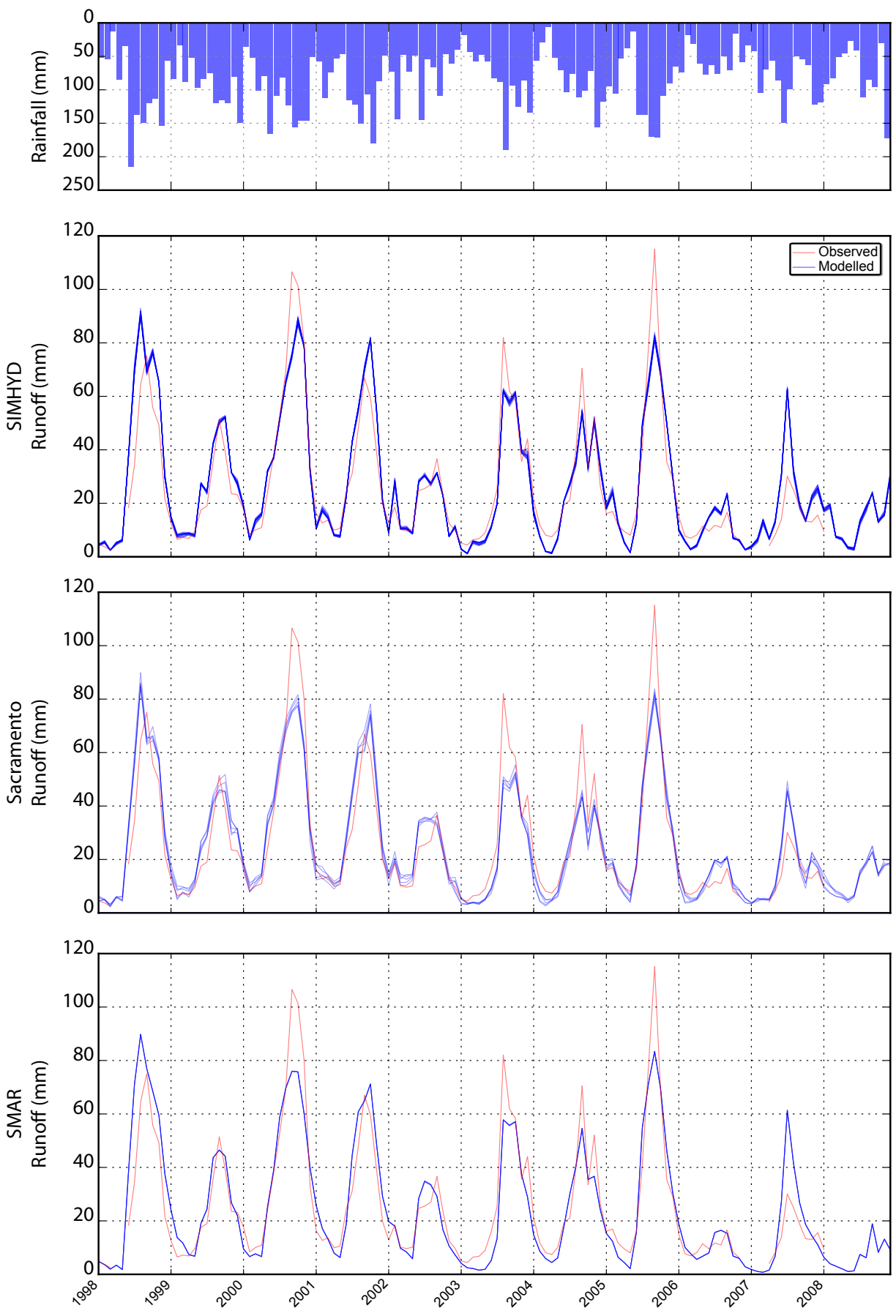
Figure 20: Monthly hydrographs of rainfall-runoff models for Biggara
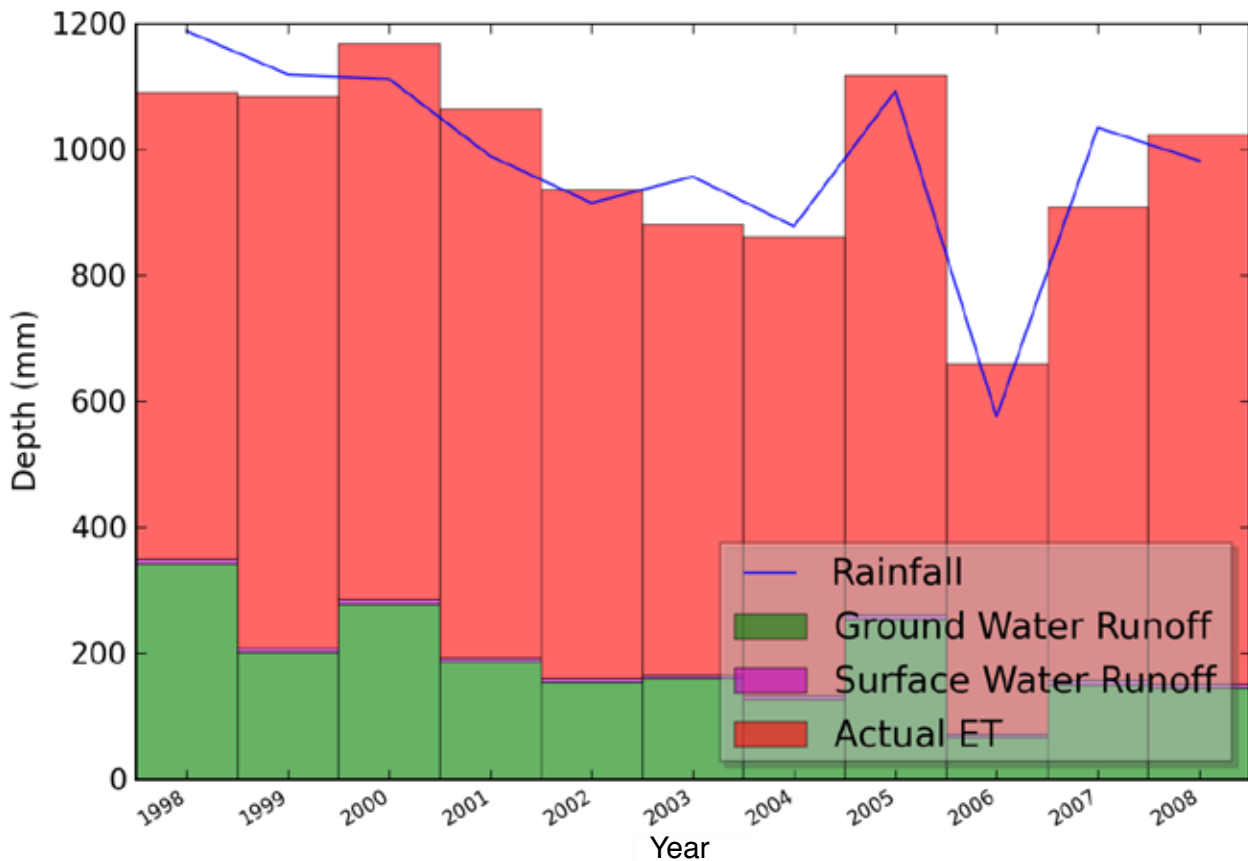
Figure 21: Yearly water balance as generated from SIMHYD for Gingera

6.2   General parameterisation versus conditional parameterisation

Skill scores of SIMHYD models calibrated using the general parameterisation scheme and the conditional parameterisation schemes are compared in Figure 22 and Figure 23. In both cases, simulations using the SIMHYD model were done with historical rainfall ensembles (section 4.2.2: past ten years of observed rainfall data) and potential evapotranspiration as inputs (Wang, Zheng & Luo 2010).

As explained in section 4.2.1, the general parameterisation scheme optimised one set of SIMHYD parameter values to the entire calibration period. In contrast, under the conditional parameterisation scheme, we have 12 sets of parameter values, each of which was the optimal parameter set for each month. As an example, the parameter set for January was obtained by optimising the objective function for January data over the calibration period. It is pointed out that each of the calibrated models was run continuously even though the model was optimised for a single month.

As shown in Figure 22, the conditional parameterisation scheme improved the accuracy of JFM forecasts (January to March) for many catchments, compared to the general parameterisation scheme. Particularly, Dohertys witnessed noticeable increase in all four accuracy indices in the first quarter. In March, RMSE increased from –9.4% to 35.3%, RMSEP from –16.4% to 24.4%, CRPS from –38.1% to 27.0% and NSE from –67.4% to 41.5% for the catchments.

For other quarters, however, most catchments experienced only marginal, if any, improvement in forecast accuracy, even though the conditional parameterisation scheme allowed more freedom in adjusting parameter values. For some cases, the monthly-varying parameter values even produced worse forecast accuracy. For instance, we found the decrease of RMSEP from 16.4% to 7.7% and from 8.9% to –8.0% in November and December respectively for Lacmalac.

In conclusion, it is questionable that the conditional parameterisation scheme is an effective solution for improving forecast accuracy enough to warrant the use of an augmented model by introducing time-varying parameter sets.
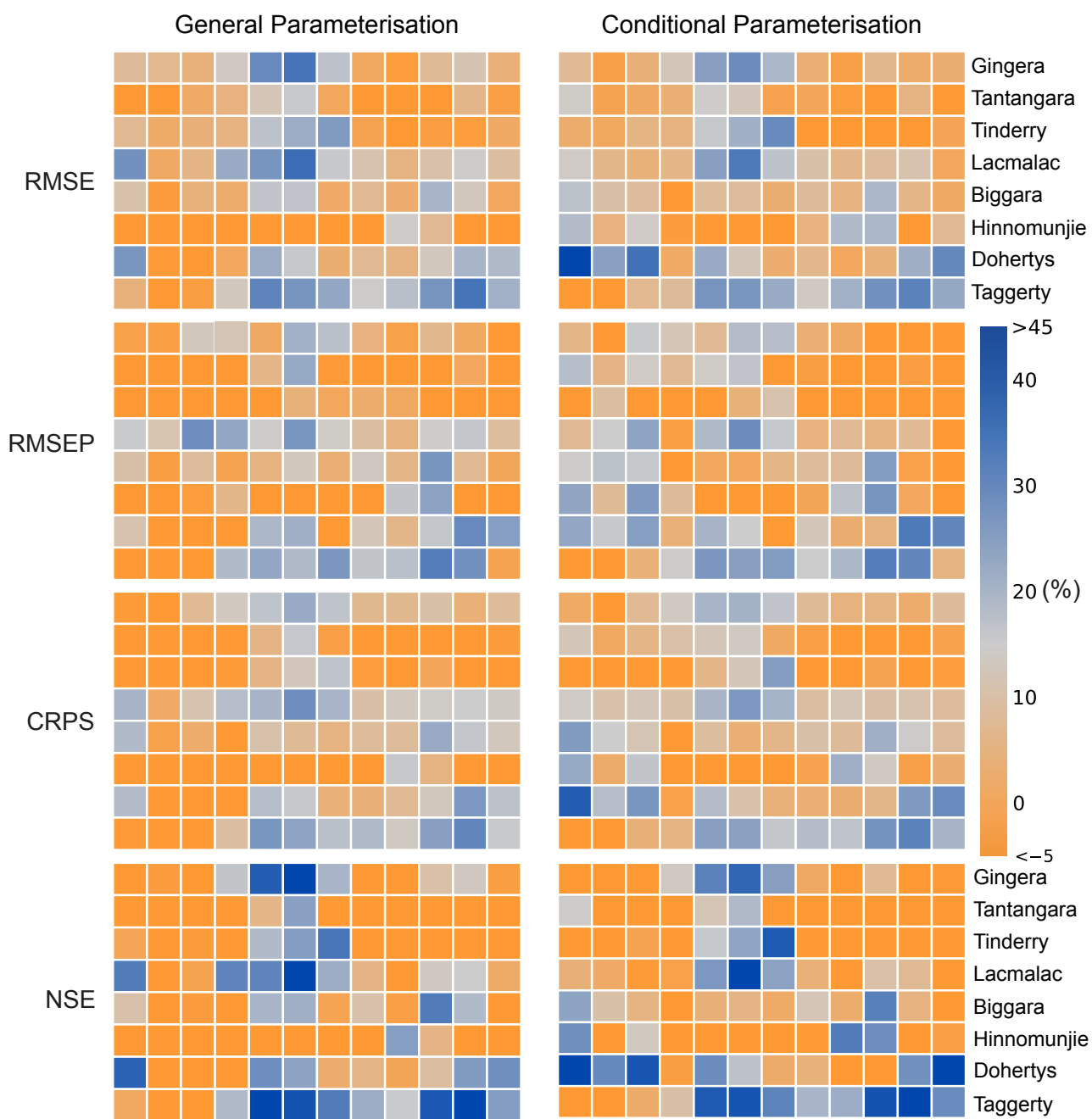
Figure 22: Skill scores, NSE and reliability of general parameterisation and conditional parameterisation. The forecasts were generated using SIMHYD with historical rainfall ensemble
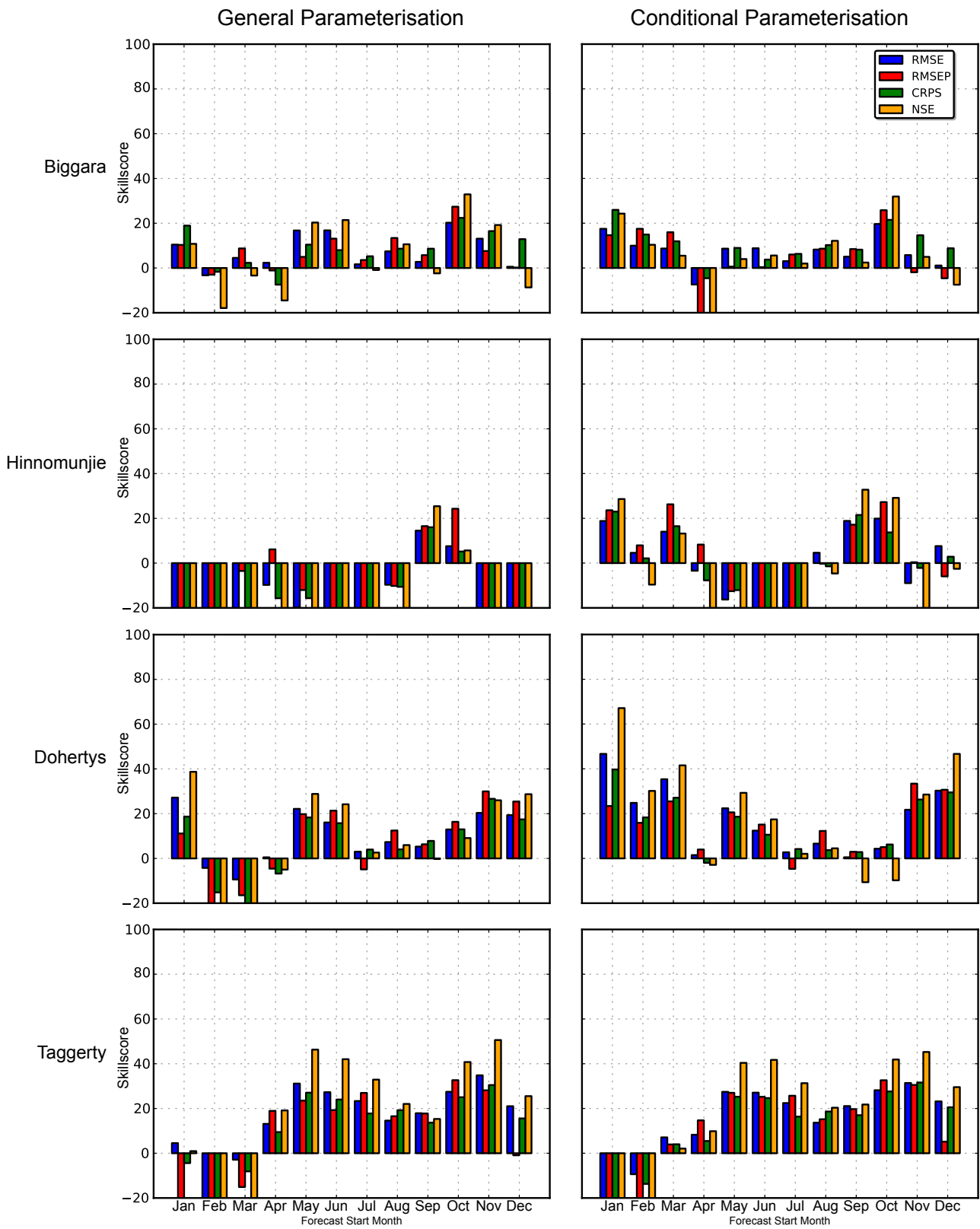
Figure 23: Bar charts for skill scores and NSE of general parameterisation and conditional parameterisation

6.3 Historical rainfall ensemble versus downscaled POAMA rainfall ensemble

Skill scores of forecasts from SIMHYD model using historical rainfall ensembles and downscaled POAMA ensembles are compared in Figure 24 and Figure 25. The SIMHYD model calibrated by the general parameterisation scheme was used for these simulations.

One recognisable feature in the right side panels was very low scores of the accuracy indices in ASON (August to November) for most catchments. For many periods, skills from downscaled POAMA ensembles declined further when compared to those from historical rainfall ensembles. For instance, Dohertys had 20.3% RMSE, 30.0% RMSEP, 26.6% CRPS and 26.0% NSE with historical rainfall ensemble for its November forecast, but those skill scores dropped sharply to 2.5%, –2.9%, –5.1% and –11.5% respectively with downscaled POAMA rainfall ensembles.

As shown in Figure 26, when we used historical rainfall ensembles, we obtained widely spread forecasts, but we could not see any noticeable bias between the medians of forecasts and observations in November. However, when we replaced the rainfall input with downscaled POAMA rainfall ensembles, in most years we had seriously underestimated forecasts compared to observations. For several years, observed streamflow went even beyond the 90th percentile of its corresponding forecasts. With downscaled POAMA rainfall ensembles, we often had systematically underestimated streamflow magnitudes particularly for high flow regimes (Figure 10).

As explained before, POAMA often underestimated rainfall magnitudes during wet seasons (Figure 9). Since most catchments of our study had high flow during ASON (see Figure 4), the underestimated rainfall forecasts resulted in lower streamflow forecasts than observation and the catchments experienced poor streamflow forecast skills during the season.

On the basis of these results, we can say that POAMA, at least version 1.5, does not demonstrate enough accuracy to use the GCM outcome, instead of a simple historical rainfall ensemble for seasonal streamflow forecast at the catchment scale. Some of the limitation may be alleviated by the improvement of downscaling methods, however fundamentally further improvements in the accuracy of POAMA rainfall forecasts are required.

While recognising that further improvements in POAMA rainfall forecasts are required in future, two important considerations are worth noting. First, rainfall forecasts and forecasts from POAMA 2.4a–c are now available, which we understand have better ocean and atmosphere assimilation techniques, and 30 ensembles members as opposed to ten from POAMA1.5 (Harry Hendon; *pers. comm.*). Second, skill scores used to quantify accuracy of streamflow forecasts rely on the difference between the median of the forecasts and streamflow observations (see equations 4 to 11). Even though downscaled POAMA rainfall and, accordingly, streamflow forecasts under prediction had poor skills, these forecasts do get the timing of highs and lows right (Figure 10). Therefore, bias in streamflow forecasts is systematic and it is possible to capture this through a posterior streamflow bias correction procedure.
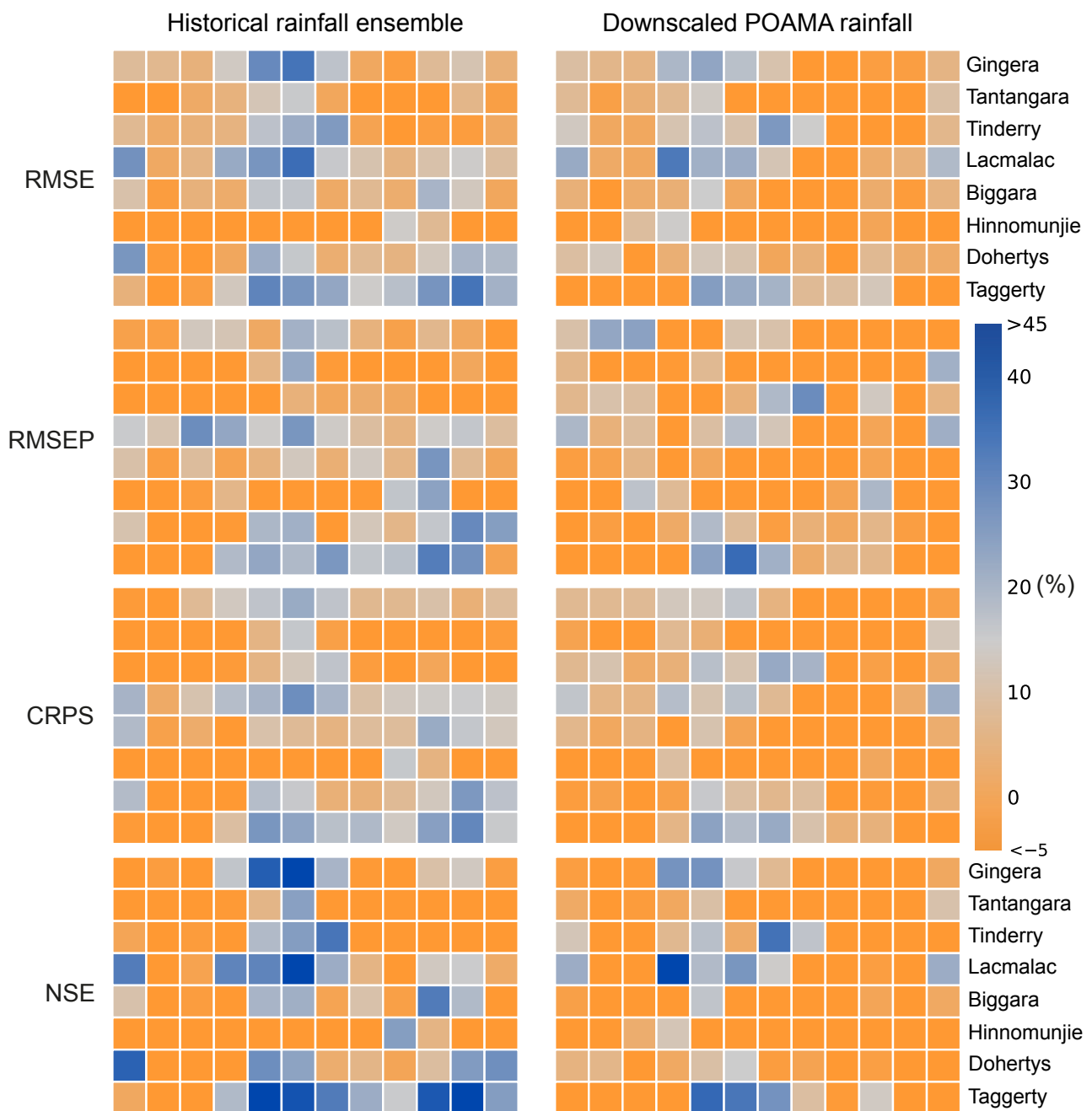
Figure 24: Skill scores, NSE and reliability of seasonal streamflow forecasts driven by historical rainfall ensemble and downscaled POAMA rainfall ensemble. The forecasts were generated using SIMHYD calibrated by general parameterisation scheme with historical rainfall ensemble
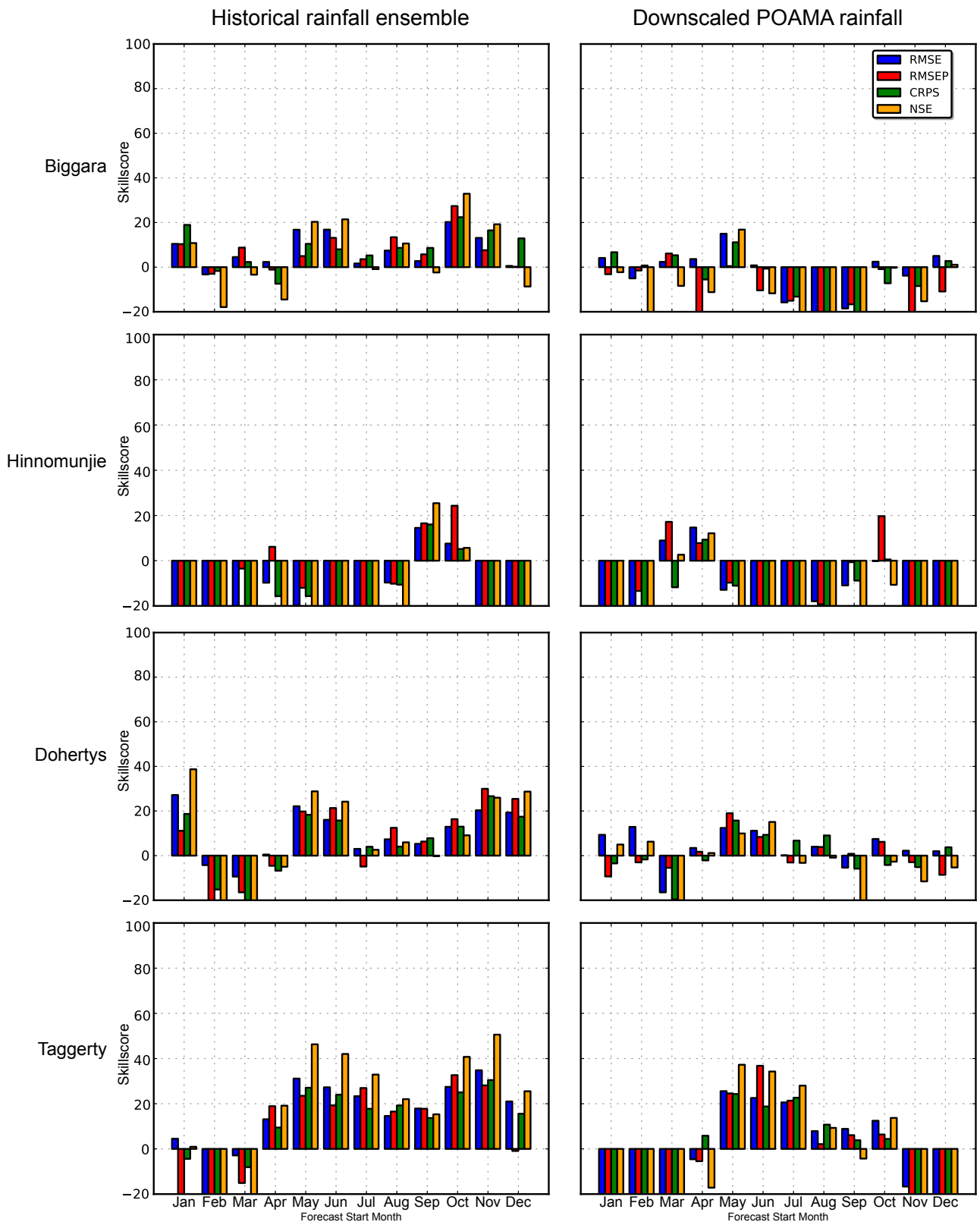
Figure 25: Bar charts for skill scores and NSE of seasonal streamflow forecasts driven by historical rainfall ensemble and downscaled POAMA rainfall ensemble
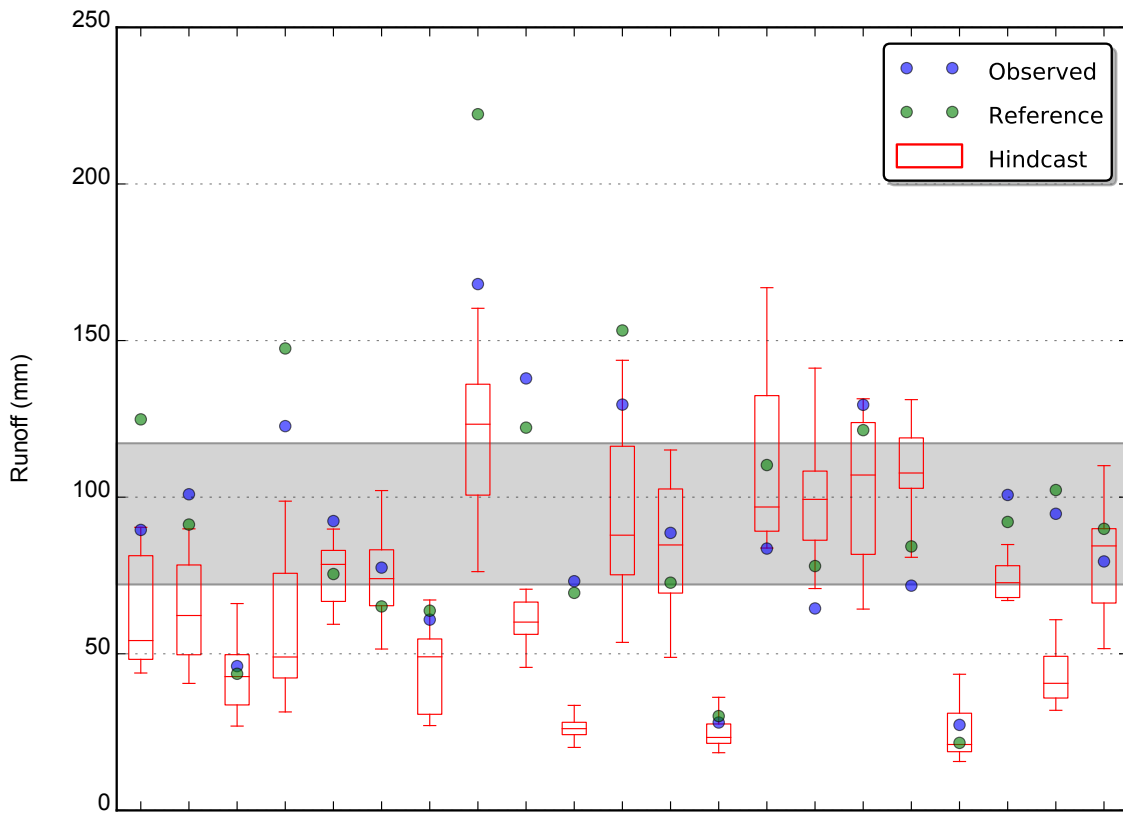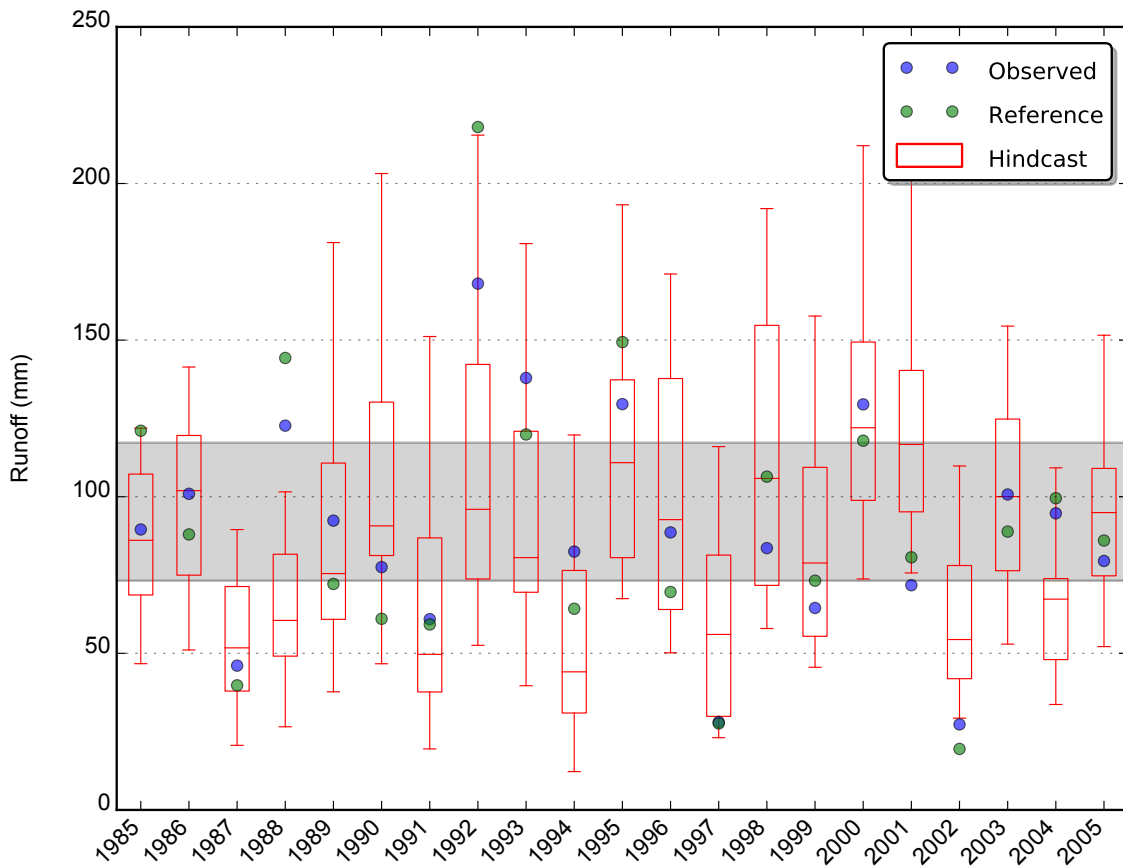
Figure 26: Monthly streamflow forecasts for Biggara in November

6.4 Posterior streamflow bias correction

Monthly rainfall and runoff dynamics of Gingera and Biggara catchments for 1995–2005 are shown in Figure 10. The hyetographs contain time series of observed and the mean of downscaled POAMA rainfall ensemble and the hydrographs contain observed streamflow (blue line), reference streamflow (observed rainfall simulated with Sacramento: green line), average of the forecast before bias correction (median of the 55 ensembles: red line) and average of the forecast after bias correction (bias corrected median of the 55 ensembles: purple line). The difference between the observed and reference flow is the bias in streamflow as a result of inaccuracies in inputs (rainfall), model structural error comprising uncertainties in derived parameters and numerical procedures, and the outputs (streamflow observations). However, the difference between reference streamflow and average of the forecast before bias correction is the bias resulting from replacing the observed rainfall with the downscaled POAMA rainfall ensembles.

In general, the forecast flow before bias correction underestimates the observed flow with much larger differences due mainly to rainfall under prediction. Note that the timing of the high and low forecast flows is aligned well and that a systematic pattern in the difference is discernible. A similar pattern was also found in other catchments. The bias correction procedure described in section 4.2.3 using the ARMA model was applied to total bias; i.e. the difference between observed streamflow (blue line) and the median of 55 forecast ensembles.

The ARMA model was calibrated using two split sampling approaches. In split sampling scheme 1, the model was calibrated to 1985–1998 (14 years) and validated against 2000–2005 (six years). In split sampling scheme 2, the respective calibration and validation periods were 1992–2005 (14 years) and 1985–1991 (seven years). Following the two split sampling approaches, the model was finally calibrated to the complete forecast dataset for 1985–2005, referred to as the combined calibration scheme.

Results from implementation of the ARMA model on target catchments at monthly time steps are summarised in Table 9, Table 10 and Table 11. In the first split sampling scheme (Table 9), ARMA correction increases the range of NSE from 0.41–0.75 to 0.59–0.85 for the calibration period and from 0.36–0.78 to 0.49–0.85 for the validation period, for all catchments except Tinderry. We could see similar improvement for other schemes.

Table 9: ARMA model parameter and NSE values of split sampling scheme 1 for monthly flow updated fortnightly

Catchment	ARMA Model	(p,q)	AR1	AR2	MA1	MA2	NSE_pc	NSE_c	NSE_pv	NSE_v
		Split sampling scheme 1: Calibration (1985–1998), Validation (2000–2005)								
Biggara	A	(1,2)	-0.84		1.33	0.36	0.62	0.81	0.60	0.64
	B	(0,1)			0.21		0.57	0.73	0.66	0.77
Dohertys	A	(1,0)	0.33				0.69	0.81	0.77	0.81
	B	(1,1)	0.77		-0.59		0.70	0.75	0.67	0.70
Gingera	A	(0,1)			0.36		0.41	0.60	0.36	0.49
	B	(0,1)			0.22		0.42	0.59	0.46	0.64
Hinnomunjie	A	(1,1)	0.16		-0.18		0.54	0.80	0.45	0.69
	B	(2,2)	0.87	-0.78	-1.03	0.78	0.48	0.73	0.45	0.71
Lacmalac	A	(1,0)	0.35				0.56	0.64	0.63	0.68
	B	(1,1)	0.82		-0.61		0.60	0.73	0.69	0.75
Taggerty	A	(1,1)	0.79		-0.60		0.75	0.85	0.78	0.85
	B	(2,0)	0.17	0.14			0.75	0.82	0.77	0.78
Tantangara	A	(0,1)			0.33		0.51	0.69	0.45	0.69
	B	(0,1)			0.16		0.55	0.72	0.60	0.77
Tinderry	A	(0,1)			0.23		0.18	0.39	-1.05	-5.07
	B	(1,1)	0.85		-0.69		0.25	0.47	-0.89	-2.30

A and B represent ARMA bias correction models implemented on the 1st and 16th day of each month respectively. NSC_pc: NSE over the calibration period prior to bias correction; NSC_c: NSE over the calibration period after bias correction; NSC_pv: NSE over the validation period prior to bias correction; NSC_v: NSE over the validation period after bias correction; (p,q): order of the ARMA model.

Table 10: ARMA model parameter and NSE values of split sampling scheme 2 for monthly flow updated fortnightly

Catchment	ARMA Model	ARMA	AR1	AR2	AR3	MA1	MA2	MA3	NSE_pc	NSE_c	NSE_pv	NSE_v
		Split sampling scheme 2: Calibration (1992–2005), Validation (1985–1990)										
Biggara	A	(1,0)	0.42						0.59	0.82	0.62	0.66
	B	(2,1)	1.08	-0.19		-0.79			0.57	0.76	0.65	0.69
Dohertys	A	(1,0)	0.41						0.70	0.83	0.73	0.75
	B	(1,1)	0.80			-0.59			0.70	0.74	0.69	0.65
Gingera	A	(1,0)	0.33						0.41	0.64	0.36	0.38
	B	(1,1)	0.82			-0.68			0.44	0.64	0.37	0.48
Hinnomunjie	A	(0,1)				0.16			0.54	0.80	0.45	0.64
	B	(0,0)							0.48	0.72	0.50	0.65
Lacmalac	A	(2,1)	1.13	-0.20		-0.79			0.59	0.73	0.51	0.46
	B	(3,3)	0.63	0.38	-0.18	-0.34	-0.50	0.29	0.64	0.75	0.57	0.59
Taggerty	A	(2,0)	0.22	0.26					0.76	0.87	0.73	0.76
	B	(1,2)	0.60			-0.47	0.24		0.76	0.81	0.75	0.79
Tantangara	A	(1,0)	0.35						0.51	0.77	0.45	0.50
	B	(2,2)	1.83	-0.84		-1.72	0.72		0.59	0.76	0.51	0.64
Tinderry	A	(1,2)	0.98			-1.06	0.11		-0.25	-0.03	0.26	0.22
	B	(0,0)							0.07	0.19	0.26	0.24

A and B represent ARMA bias correction models implemented on the 1st and 16th day of each month respectively. NSC_pc: NSE over the calibration period prior to bias correction; NSC_c: NSE over the calibration period after bias correction; NSC_pv: NSE over the validation period prior to bias correction; NSC_v: NSE over the validation period after bias correction; (p,q): order of the ARMA model.

Table 11: ARMA model parameter and NSE values of the combined calibration scheme for monthly flow updated fortnightly

| Catchment | ARMA Model | Combined calibration scheme: Calibration (1985–2005), Validation (–) | | | | | | | | |
		ARMA	AR1	AR2	MA1	MA2	NSE_pc	NSE_c	NSE_pv	NSE_v
Biggara	A	(0,1)			0.38		0.62	0.79		
	B	(1,0)	0.22				0.60	0.75		
Dohertys	A	(1,0)	0.30				0.71	0.82		
	B	(1,1)	0.78		-0.62		0.70	0.74		
Gingera	A	(1,2)	0.97		-0.65	-0.25	0.43	0.61		
	B	(1,2)	0.96		-0.74	-0.16	0.45	0.63		
Hinnomunjie	A	(1,1)	-0.84		0.99		0.52	0.78		
	B	(0,0)					0.48	0.72		
Lacmalac	A	(1,2)	0.90		-0.62	-0.13	0.59	0.67		
	B	(1,1)	0.87		-0.69		0.63	0.74		
Taggerty	A	(1,2)	0.63		-0.52	0.15	0.76	0.85		
	B	(1,2)	0.56		-0.43	0.12	0.76	0.82		
Tantangara	A	(0,1)			0.24		0.51	0.71		
	B	(0,1)			0.12		0.57	0.74		
Tinderry	A	(1,2)	0.98		-0.75	-0.18	0.22	0.44		
	B	(1,1)	0.91		-0.76		0.29	0.48		

A and B represent ARMA bias correction models implemented on the 1st and 16th day of each month respectively. NSC_pc: NSE over the calibration period prior to bias correction; NSC_c: NSE over the calibration period after bias correction; NSC_pv: NSE over the validation period prior to bias correction; NSC_v: NSE over the validation period after bias correction ; (p,q): order of the ARMA model.

Similarly, significant improvements were also achieved at three-monthly time steps (Table 12, Table 13 and Table 14). In the same first split sampling scheme (Table 12), pre-ARMA bias correction, NSE values over the calibration and validation periods vary in the range 0.12 to 0.72 and –0.04 to 0.65 respectively. The only exception is Tinderry where NSE values for the calibration and validation periods were –0.19 and –1.81 respectively, indicating poor forecasts relative to those from reference climatology. NSE values of post-ARMA bias correction over the calibration and validation periods vary in the range 0.34 to 0.85 and 0.27 to 0.9 respectively, indicating significant improvement in streamflow forecasts.

Table 12: ARMA model parameter and NSE values of split sampling scheme 1 for three-monthly flow updated monthly

Catchment	Model	(p,q)	AR1	AR2	MA1	MA2	NSE_pc	NSE_c	NSE_pv	NSE_v
						Split sampling scheme 1: Calibration (1985–1998), Validation (2000–2005)				
Biggara	A	(0,1)			0.39		0.59	0.82	0.42	0.61
	B	(1,0)	0.23				0.51	0.66	0.32	0.74
	C	(0,0)					0.43	0.65	0.50	0.82
Dohertys	A	(0,0)					0.72	0.82	0.51	0.57
	B	(0,0)					0.67	0.76	0.64	0.85
	C	(0,1)			0.21		0.59	0.75	0.65	0.86
Gingera	A	(0,0)					0.38	0.57	0.14	0.29
	B	(0,1)			0.23		0.18	0.39	-0.04	0.27
	C	(0,0)					0.12	0.34	0.07	0.47
Hinnomunjie	A	(0,0)					0.48	0.77	0.36	0.67
	B	(0,0)					0.59	0.77	0.27	0.73
	C	(0,0)					0.43	0.76	0.34	0.85
Lacmalac	A	(1,0)	0.25				0.62	0.80	0.54	0.71
	B	(1,0)	0.26				0.36	0.44	0.37	0.50
	C	(0,1)			0.34		0.41	0.58	0.58	0.78
Taggerty	A	(1,0)	0.26				0.70	0.85	0.55	0.69
	B	(1,0)	0.25				0.66	0.79	0.53	0.81
	C	(0,1)			0.35		0.57	0.76	0.64	0.78
Tantangara	A	(0,0)					0.58	0.79	0.43	0.76
	B	(1,0)	0.21				0.41	0.61	0.21	0.66
	C	(1,1)	-0.54		0.85		0.40	0.72	0.43	0.90
Tinderry	A	(1,0)	0.46				0.04	0.45	-1.73	-5.52
	B	(0,1)			0.33		-0.03	0.30	-0.65	-10.05
	C	(1,0)	0.25				-0.19	0.14	-1.81	-12.27

A, B and C represent ARMA bias correction models implemented on the 1st day of the following months: Model A: Jan, Apr, Jul and Oct, Model B: Feb, May, Aug, Nov; and Model C: Mar, Jun, Sep and Dec. NSC_pc: NSE over the calibration period prior to bias correction; NSC_c: NSE over the calibration period after bias correction; NSC_pv: NSE over the validation period prior to bias correction; NSC_v: NSE over the validation period after bias correction; (p,q): order of the ARMA model.

Table 13: ARMA model parameter and NSE values of split sampling scheme 2 for three-monthly flow updated monthly

Catchment	Model	(p,q)	AR1	AR2	MA1	MA2	NSE_pc	NSE_c	NSE_pv	NSE_v
		Split sampling scheme 2: Calibration (1992–2005), Validation (1985–1990)								
Biggara	A	(0,1)			0.45		0.51	0.72	0.64	0.82
	B	(1,0)	0.28				0.40	0.69	0.62	0.58
	C	(1,0)	0.33				0.42	0.67	0.53	0.72
Dohertys	A	(1,0)	0.26				0.63	0.70	0.81	0.88
	B	(0,0)					0.66	0.78	0.74	0.78
	C	(1,0)	0.40				0.60	0.77	0.68	0.76
Gingera	A	(1,0)	0.22				0.30	0.50	0.41	0.50
	B	(1,1)	0.16		0.31		0.28	0.60	0.01	-0.38
	C	(1,0)	0.21				0.13	0.40	0.12	0.22
Hinnomunjie	A	(0,1)			0.39		0.48	0.68	0.39	0.68
	B	(0,0)					0.49	0.81	0.56	0.57
	C	(1,0)	0.36				0.43	0.81	0.31	0.67
Lacmalac	A	(1,0)	0.38				0.55	0.72	0.66	0.81
	B	(2,0)	0.16	0.28			0.38	0.49	0.31	0.20
	C	(1,0)	0.34				0.48	0.58	0.37	0.42
Taggerty	A	(1,0)	0.43				0.64	0.79	0.77	0.87
	B	(0,1)			0.51		0.65	0.81	0.64	0.49
	C	(0,1)			0.48		0.58	0.79	0.60	0.64
Tantangara	A	(1,0)	0.30				0.51	0.74	0.60	0.84
	B	(1,1)	0.79		-0.63		0.35	0.68	0.43	0.39
	C	(1,0)	0.28				0.43	0.74	0.37	0.58
Tinderry	A	(1,0)	0.31				0.11	0.25	-0.03	0.18
	B	(0,0)					0.21	0.25	-0.04	-0.02
	C	(1,1)	-0.81		0.99		-0.50	-0.33	-0.28	-0.25

A, B and C represent ARMA bias correction models implemented on the 1st day of the following months: Model A: Jan, Apr, Jul and Oct, Model B: Feb, May, Aug, Nov; and Model C: Mar, Jun, Sep and Dec. NSC_pc: NSE over the calibration period prior to bias correction; NSC_c: NSE over the calibration period after bias correction; NSC_pv: NSE over the validation period prior to bias correction; NSC_v: NSE over the validation period after bias correction; (p,q): order of the ARMA model.

Table 14: ARMA model parameter and NSE values of the combined calibration scheme for three-monthly flow updated monthly

		Combined calibration scheme: Calibration (1985–2005), Validation (–)								
Catchment	Model	(p,q)	AR1	AR2	MA1	MA2	NSE_pc	NSE_c	NSE_pv	NSE_v
Biggara	A	(0,1)			0.36		0.56	0.77		
	B	(1,0)	0.22				0.47	0.69		
	C	(0,1)			0.19		0.46	0.69		
Dohertys	A	(0,0)					0.69	0.78		
	B	(0,0)					0.67	0.78		
	C	(2,2)	0.18	0.65	0.15	-0.76	0.62	0.80		
Gingera	A	(1,1)	0.92		-0.80		0.39	0.60		
	B	(1,0)	0.22				0.20	0.43		
	C	(1,1)	-0.60		0.96		0.16	0.51		
Hinnomunjie	A	(0,0)					0.46	0.74		
	B	(0,0)					0.51	0.77		
	C	(1,0)	0.21				0.41	0.79		
Lacmalac	A	(1,1)	0.81		-0.56		0.62	0.79		
	B	(1,0)	0.24				0.40	0.51		
	C	(1,1)	-0.52		0.90		0.47	0.70		
Taggerty	A	(1,0)	0.30				0.69	0.82		
	B	(1,0)	0.26				0.65	0.79		
	C	(0,1)			0.31		0.60	0.77		
Tantangara	A	(1,0)	0.20				0.55	0.79		
	B	(1,0)	0.16				0.37	0.65		
	C	(1,1)	-0.55		0.87		0.42	0.77		
Tinderry	A	(1,1)	0.84		-0.50		0.14	0.50		
	B	(0,1)			0.27		0.08	0.30		
	C	(1,1)	0.86		-0.67		-0.07	0.18		

A, B and C represent ARMA bias correction models implemented on the 1st day of the following months: Model A: Jan, Apr, Jul and Oct, Model B: Feb, May, Aug, Nov; and Model C: Mar, Jun, Sep and Dec. NSC_pc: NSE over the calibration period prior to bias correction; NSC_c: NSE over the calibration period after bias correction; NSC_pv: NSE over the validation period prior to bias correction; NSC_v: NSE over the validation period after bias correction ; (p,q): order of the ARMA model.

Forecasting results for the split sampling scheme 1 pre- and post-ARMA bias correction are shown as scatter plots in Figure 27 and Figure 28 respectively. Significant improvements in streamflow forecasts through the bias correction procedure are noticeable here. Note that observed streamflow over the validation period (2000–2005) is about 50–70% of the range over the calibration period (1985–1998) indicating a very dry validation period. This difference has significant impact on the mean and standard deviation of the monthly (and three-monthly) streamflow bias over the two periods.

It is pointed out that during standardisation and de-standardisation of the streamflow bias time series (steps 3 and 6, section 4.2.3), mean and standard deviations over the calibration period alone were used for the validation period also. Noting that since the model validation period was very dry, the model was therefore tested on a more extreme situation than is likely in the operational service, whereas the ARMA model would be calibrated to all observed data prior to performing streamflow bias correction in the forecast mode.
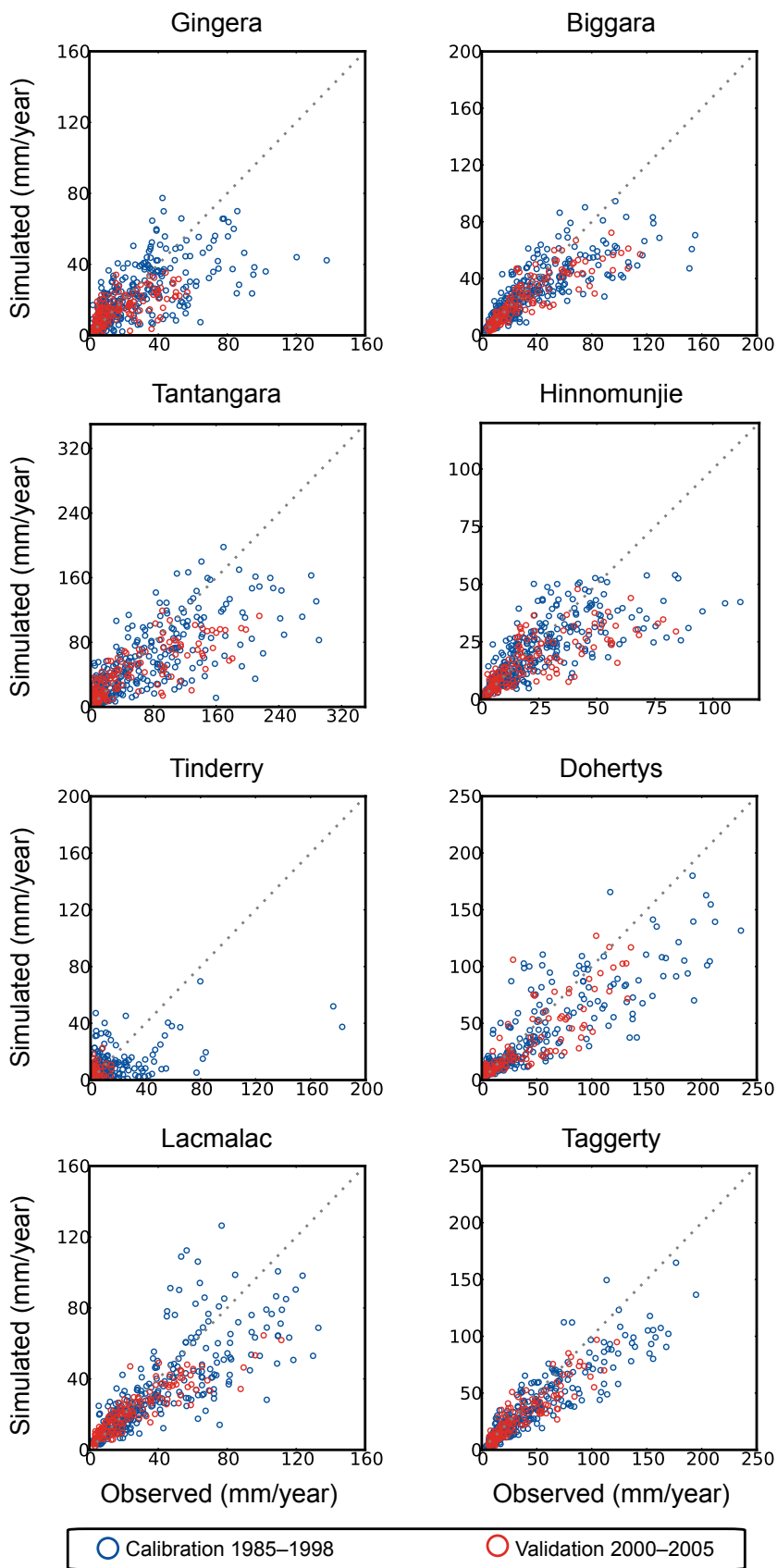
Figure 27: Observed and median monthly streamflow forecasts prior to bias correction for the calibration and validation periods. Note that the calibration (1985–1998) and validation (2000–2005) periods refer to split sampling scheme used for implementing the ARMA mode
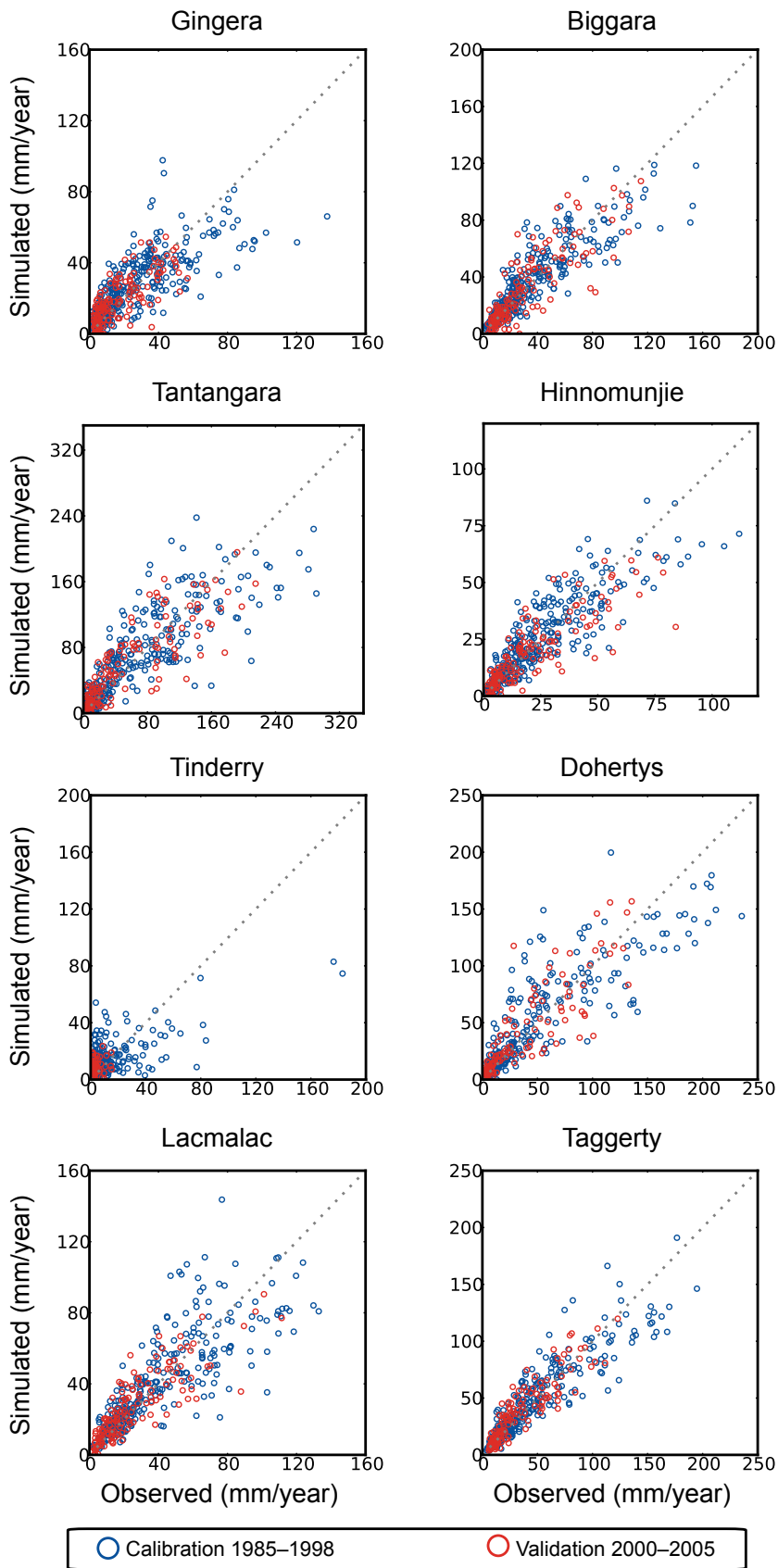
Figure 28: Observed and bias corrected median monthly streamflow forecasts for the calibration and validation periods. Note that the calibration (1985–1998) and validation (2000–2005) periods refer to split sampling scheme used for implementing the ARMA model

Results from the second split sampling approach indicated similar significant improvements at monthly as well as at three-monthly time steps (Table 10 and Table 13). Here the model is calibrated to relatively dry conditions (1990–2005) and validated on moderate to wet conditions (1985–1990).

After validating the model over two split sampling schemes, the ARMA model was calibrated using the combined calibration scheme to the complete forecast period 1985 to 2005 on all target catchments. The ACF and PACF of the monthly streamflow bias (X_t = observed minus median forecast pre-ARMA bias correction), standardised streamflow bias (y_t)

and the residual series ($\hat{\varepsilon}_t$ = remaining residuals or the white noise post-ARMA bias correction) for the Gingera and Biggara catchments are shown in Figure 29 and Figure 30 respectively. The ACF and PACF of the monthly streamflow bias (X_t) and standardised streamflow bias (y_t) are indicative of the persistence captured by the ARMA model. However, as shown in PACF (one month lag) there is no significant persistence left in the residuals ($\hat{\varepsilon}_t$) from the ARMA model following implementation of the bias correction procedure.
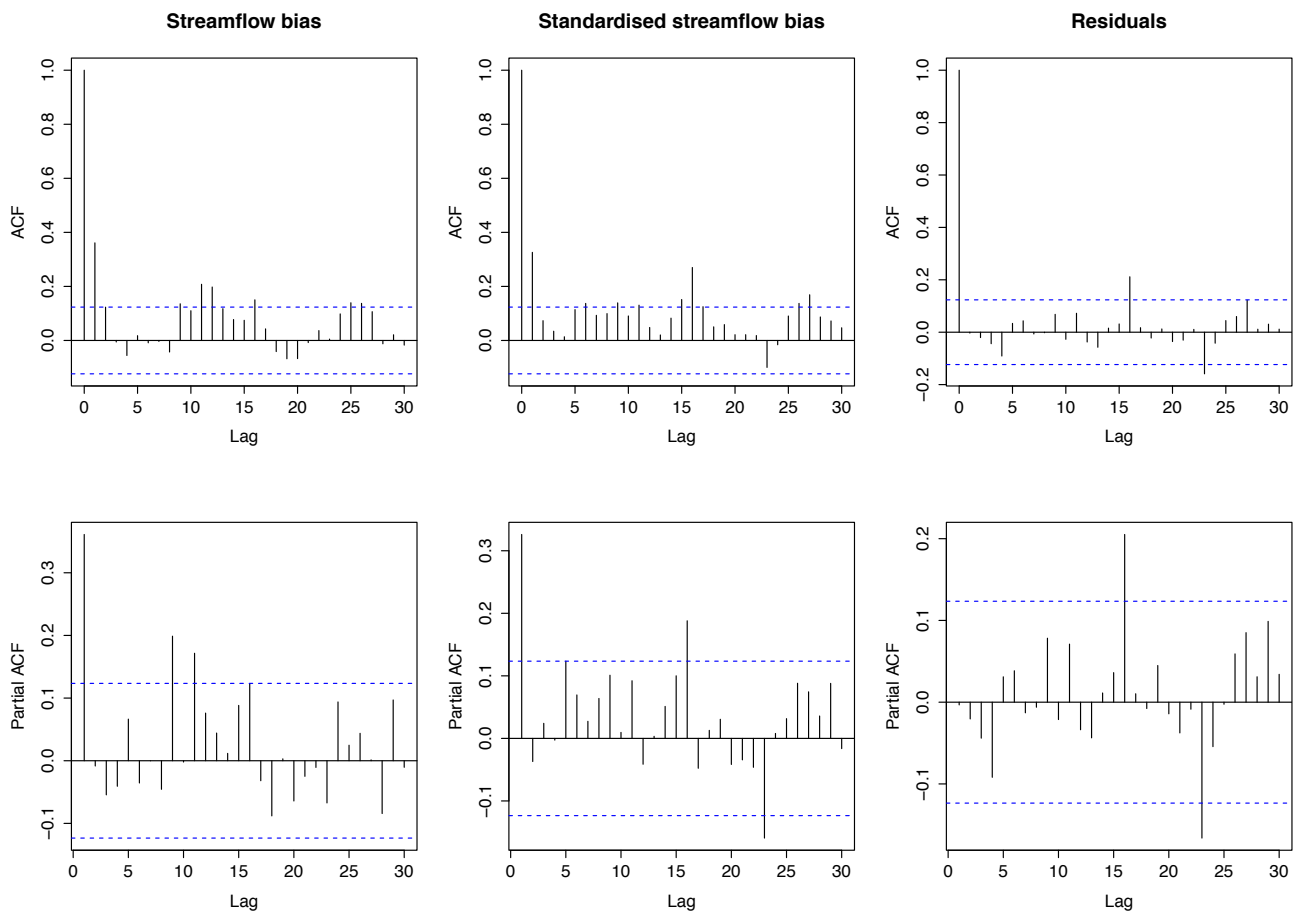
Figure 29: Autocorrelation function of residuals of Gingera before and after ARMA-based bias correction
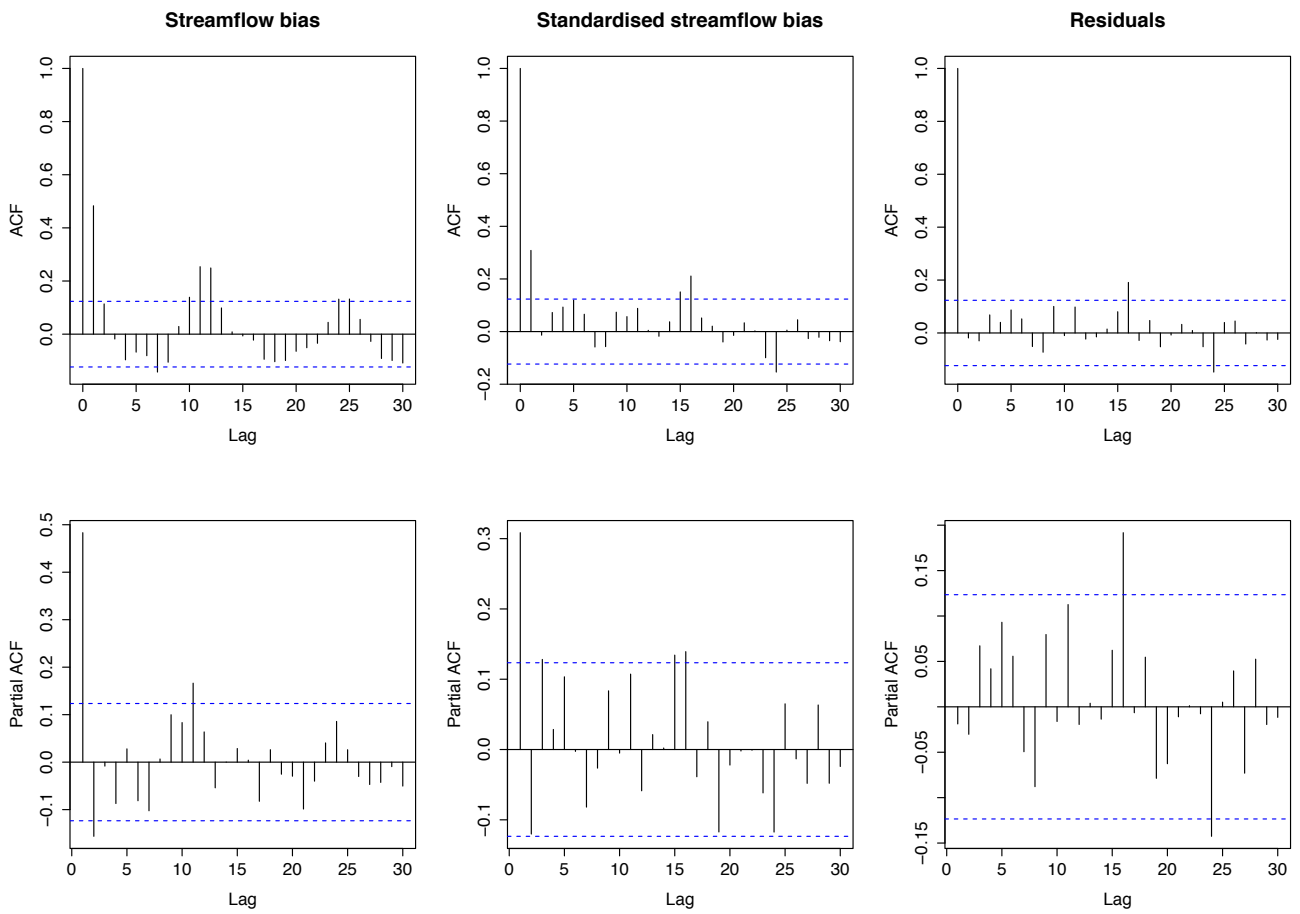
Figure 30: Autocorrelation function of residuals of Biggara before and after ARMA-based bias correction

Skill scores of three-monthly streamflow forecasts pre- and post-ARMA-based bias correction procedure for all catchments are summarised in Figure 31. The comparison shows marked noticeable improvement in skill scores of the forecasts post-ARMA-based bias correction. Reliability diagrams (predictive QQ plots of PIT plots) for each month for Biggara and Gingera catchments (Figure 32 and Figure 33) demonstrate that the bias correction produced quite reliable streamflow forecast distributions compared to observed data. It is worth pointing out that the bias correction procedure was implemented on time series of the median monthly or three-monthly biased streamflow forecasts i.e. median of the 55 streamflow ensembles at every time step of the forecast period. The correction at each time step was then imposed on each member of the set of 55 ensembles along with the random term ε_t (equation 11, Appendix A). Therefore, the entire forecast distribution at each time step was corrected for bias in streamflow forecast. We believe that it is for this reason that the bias corrected streamflow forecasts are accurate and reliable.

Note that the automated ARMA model calibration scheme of Hyndman and Khandakar (2008) adopted in this study guarantees that the fitted models are stationary and invertible, i.e. roots of the equation $\Phi(B) = 0$ and $\Pi(B) = 0$ or zeroes of the polynomial $\Phi(B)$ and $\Pi(B)$ lie outside the unit circle (see section 4.2.3). This issue is particularly important for the Bureau because the seasonal streamflow forecasting service will require a high level of automation across many sites nationally into the future.

It can be concluded that there is a very strong case for the use of an ARMA-based bias correction procedure in dynamic seasonal streamflow forecasting and that a catchment specific ARMA model fitted across all update time steps will offer significant improvements in the streamflow forecasts. Note that two separate ARMA models are required for fortnightly updates of monthly flow and three separate ARMA models are required for monthly updates of three-monthly flow forecasts to avoid data overlap, i.e. more frequent updates at sub-monthly time steps makes model form less parsimonious. We can also calibrate the ARMA model with recent data, which will make our bias correction more adaptive to recently observed climate condition. The adaptive application of ARMA model for bias correction will be explored in future work.
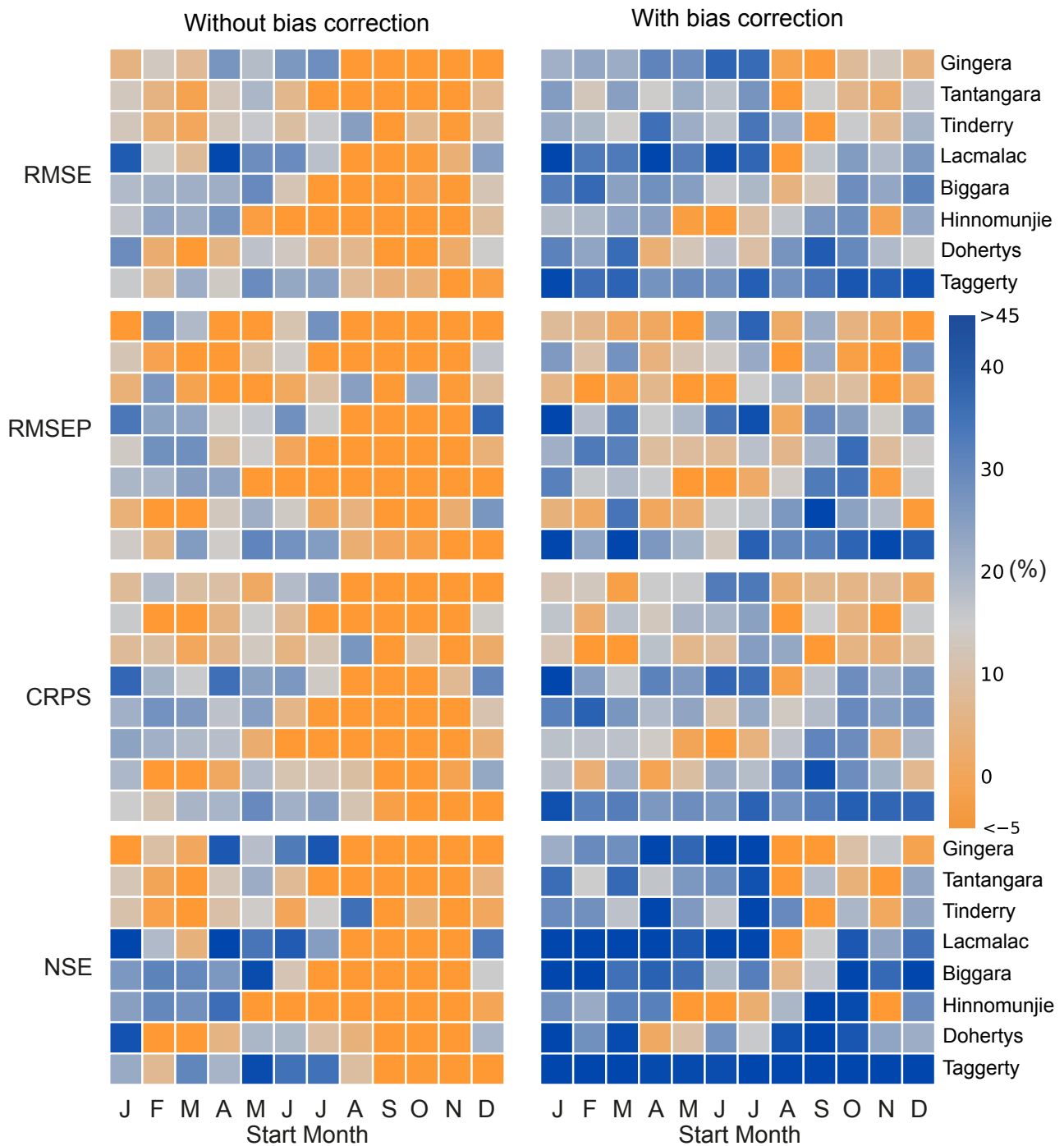
Figure 31: Skill scores, NSE and reliability of seasonal streamflow forecasts before and after ARMA-based bias correction. The forecasts were generated using Sacramento calibrated with downscaled POAMA rainfall ensemble
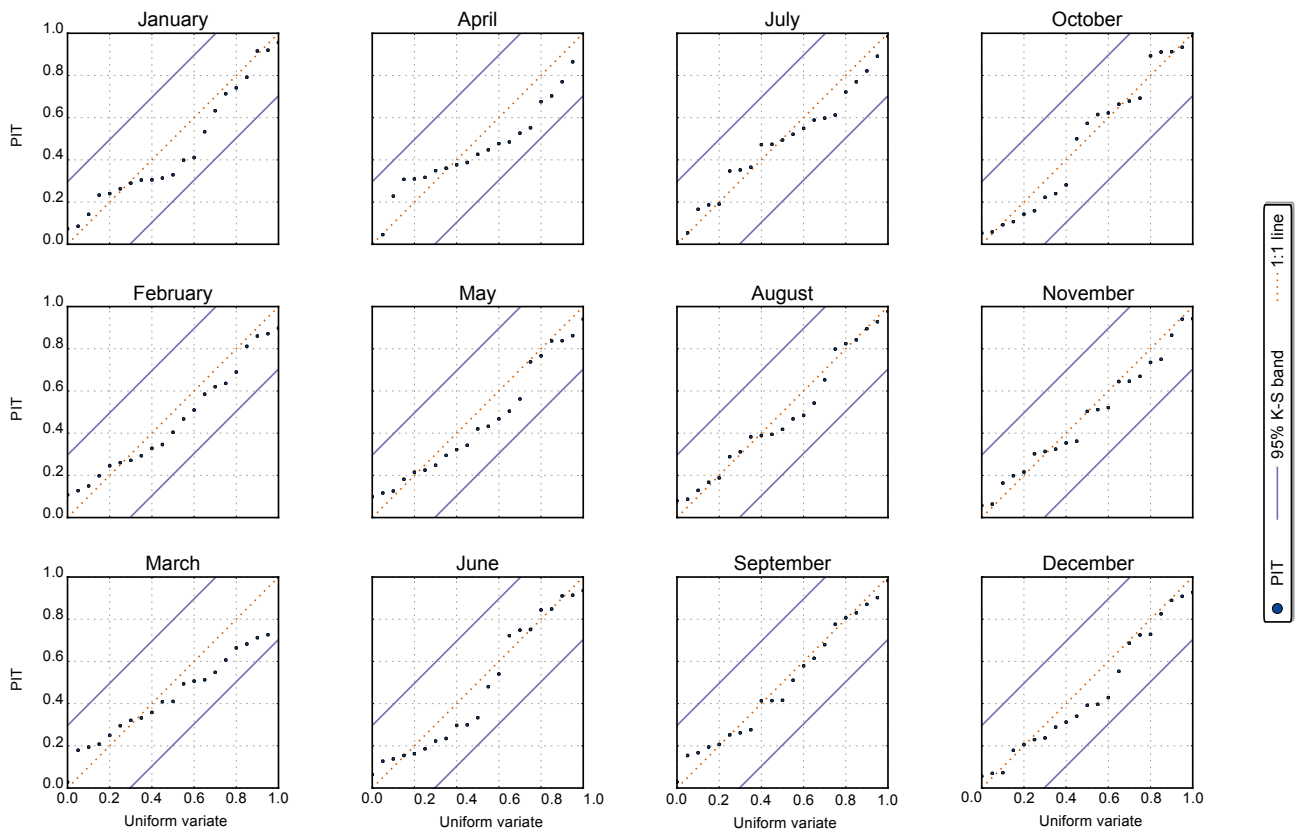
Figure 32: Predictive QQ plot of PITs from seasonal streamflow forecasts for Biggara
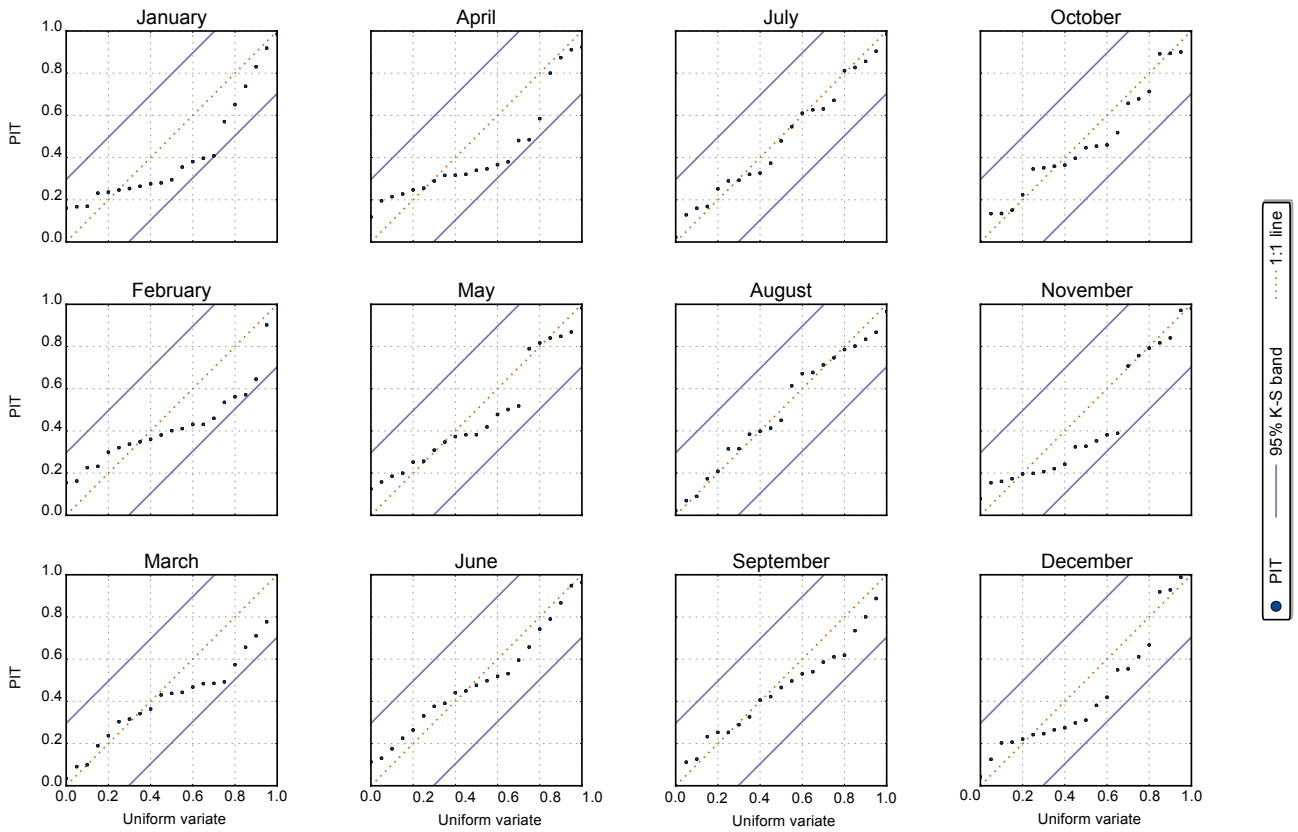
Figure 33: Predictive QQ plot of PITs from seasonal streamflow forecasts for Gingera

6.5   Dynamic versus BJP

In this subsection, we compare results from the dynamic modelling approach with those from the BJP, which is the statistical modelling approach currently used for the operational Seasonal Streamflow Forecasting (SSF) products. Skill scores of three-monthly streamflow forecasts from the Sacramento model simulated with downscaled POAMA rainfall with those derived from the operational statistical BJP model of Wang and Robertson (2011) are compared in Figure 34 and Figure 35. Figure 34 compares forecasts from Sacramento before ARMA bias correction while Figure 35 shows comparison after bias correction. It is pointed out that skill scores for the Tantangara Reservoir from BJP forecasts are missing because the SSF service is currently not delivered for the catchment.

Prior to the bias correction, BJP produced much better forecasts than the dynamic model (Figure 34). Interestingly, BJP showed lower skill scores during dry seasons while Sacramento produced the lowest scores during wet seasons in most catchments. Typically, the five months of least streamflow in these catchments are January to May, accounting for about 5–8% of the total mean annual streamflow (Figure 4). Further, mean monthly rainfall is relatively low during December to March and potential evapotranspiration demand is high during September to March (Figure 5 and Figure 6). At the dry end, downscaled POAMA rainfall forecasts are evenly spread across the 1:1 line (Figure 9). Noting that internal storages in the model are re-initialised with observed climatology at the time of forecast update (section 4.2.2) and that these storages will be generally low during dry periods, runoff contributions would typically occur through the groundwater flux term which, in such catchments, behaves like a damped exponential distribution. This is captured well in the dynamic modelling approach because of the longer memory of the groundwater response function than that in the statistical approach where memory is often captured through past streamflow over one to four months.
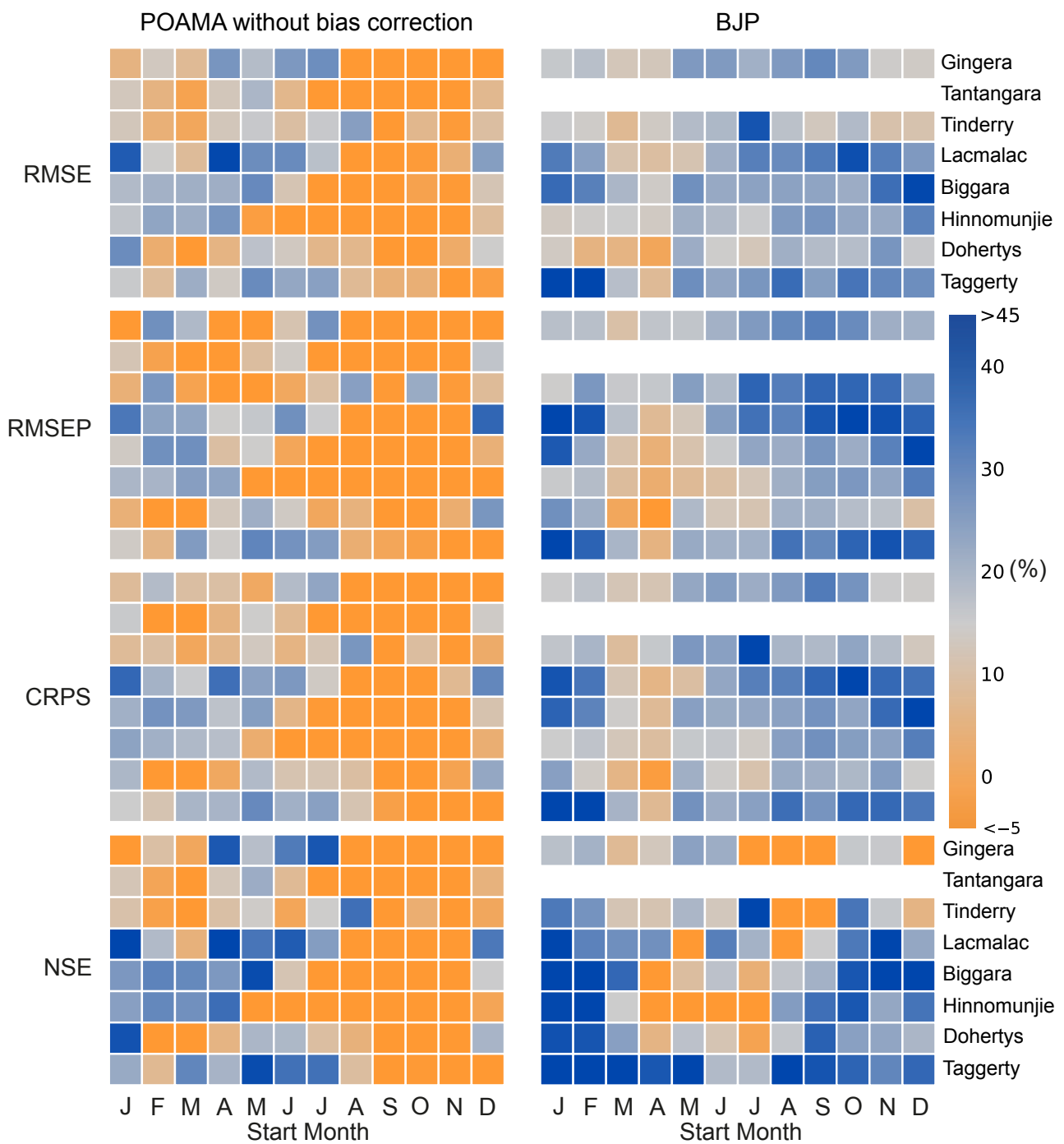
Figure 34: Skill scores, NSE and reliability of seasonal streamflow forecasts from Sacramento and BJP
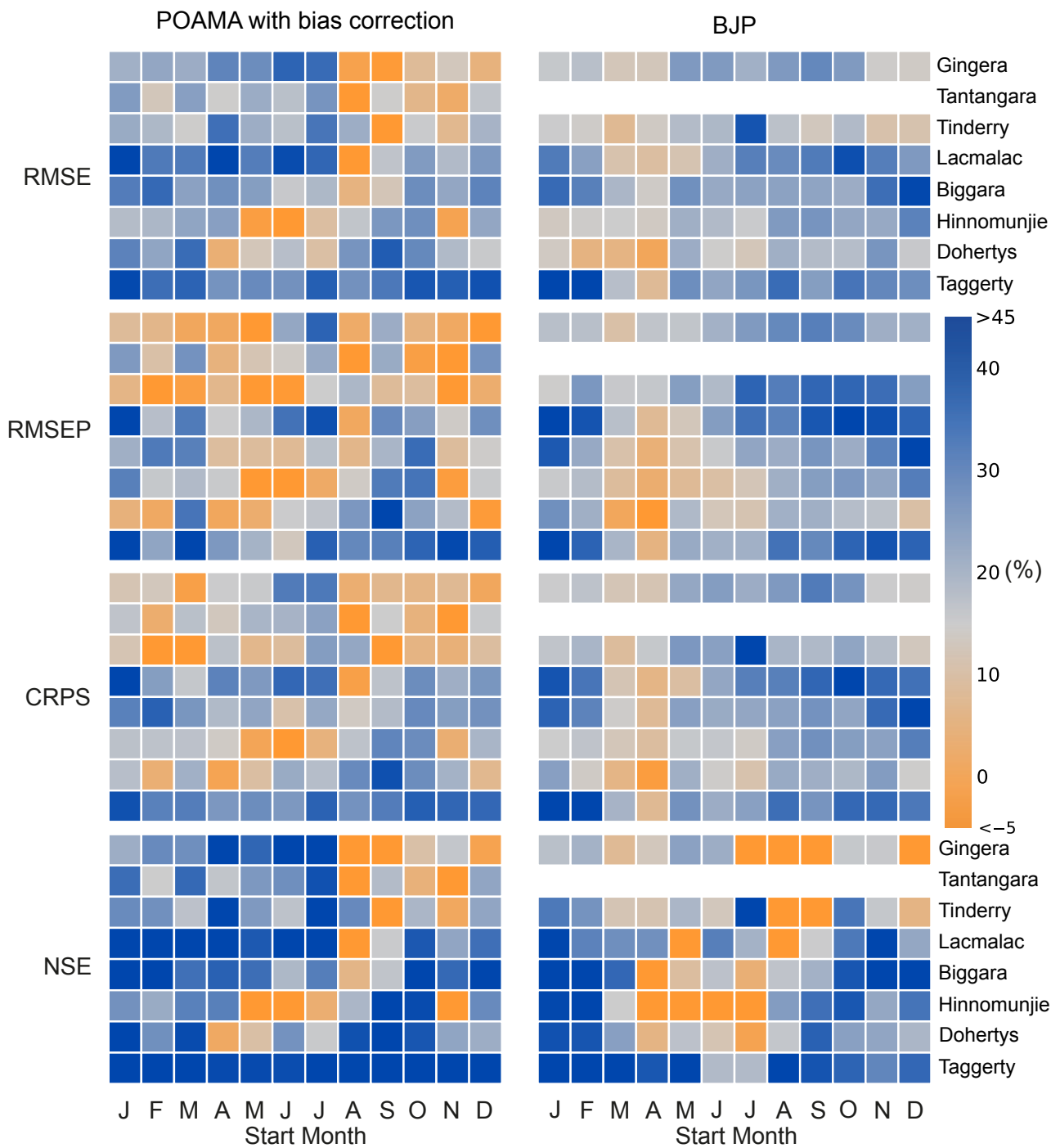
Figure 35: Skill scores, NSE and reliability of seasonal streamflow forecasts from Sacramento and BJP. The Sacramento was calibrated by the general parameterisation scheme. Its forecasts were generated with downscaled POAMA rainfall ensemble and then bias-corrected using ARMA method

The BJP approach transforms a set of streamflows and their predictors (past streamflows and climate indices) into a multivariate normal distribution and infers the distribution of model parameters of a linear regression model using a Bayesian formulation, which is implemented through a Markov Chain Monte Carlo (MCMC) sampling method. Further, a separate model is calibrated directly in the 'forecast simulation mode' to observed three-month streamflow for each monthly update time step through a cross-validation procedure (i.e. leave one year out method). This approach is different to the dynamic approach where a single hydrologic model is calibrated to historical data in retrospective mode and 'not' in the forecast mode. Note that inclusion of the bias correction procedure in the dynamic approach where bias in streamflow forecast distribution is calibrated to observed streamflow in conjunction with the hydrologic model and downscaled POAMA rainfall forecasts, is analogous to the BJP approach in that forecast distributions from both approaches are now calibrated to observed data. Therefore, it is logical to compare accuracy and reliability of the bias corrected streamflow forecasts against those from BJP rather than those prior to bias correction. Given that the dynamic approach makes use of a single hydrologic model and three ARMA models (one each for January–April–July–October, February–May–August–November and March–June–September–December to avoid data overlap of three-monthly flow at monthly update time steps), the overall model structure is more parsimonious than the 12 BJP models with each model corresponding to each monthly update time step.

Once the biases in the Sacramento streamflow outcome were removed using the ARMA model, the dynamic model turned out to produce highly skilled forecasts that are comparable or better than those from the BJP model (Figure 35). Even though the comparison draws different results depending on sites and months, overall, the dynamic approach tends to outperform BJP during January–July. However, BJP forecasts are more skilful during August–December, particularly for Gingera and Tinderry while skills are comparable in other catchments. Since the two modelling approaches seem to complement each other, we believe that streamflow forecasts blended from a combination of statistical and dynamic approaches in future work are likely to provide better streamflow forecasts. This issue will be a major topic for future WIRADA research in seasonal streamflow forecasting; using the Bayesian model averaging procedure for improved rainfall forecasting would be a good starting point for this research.

It is interesting to note the differences in skill scores for a given forecast distribution across all catchments. Results from the dynamic approach show that both RMSE and NSE, which measure accuracy in the flow domain, have high skill scores across all catchments relative to RMSEP and CRPS, which measure accuracy in the probability domain. Note that RMSE and NSE quantify skills directly in the measurement space (i.e. flow domain) and are potentially sensitive to few large errors. In contrast, RMSEP and CRPS quantify skills in the probability domain and these skill scores impose less penalty to inaccurate high streamflow forecasts. However, in the case of narrow streamflow distributions, e.g. the January–May period in most catchments (see Figure 4), small errors around median flows can be amplified greatly in the probability domain even though errors might be very small in the flow domain. Since these skill scores measure different aspects of forecast accuracy, it is therefore desirable to consider all skill scores in assessing forecast accuracy.

We believe that although downscaled POAMA rainfall and accordingly streamflow forecasts under predict observed streamflow resulting in poor skill scores prior to bias correction, these forecasts do get the timing of highs and lows right (Figure 10). Therefore, bias in streamflow forecasts is systematic and it is possible to capture this through a posterior streamflow bias correction procedure well. Since median of the streamflow forecasts is corrected well through the ARMA-based bias correction procedure, substantial improvements in the skill scores is a logical outcome as they largely quantify accuracy of streamflow forecasts based on the difference between the median of the forecasts and streamflow observations (see equations 4 to 11).

Reliability of bias corrected streamflow forecasts from the dynamic approach is assessed through predictive quantile-quantile (QQ) plots or the PIT plots (Figure 32 and Figure 33). These reliability diagrams for Biggara and Gingera show that 'three-monthly' streamflow forecasts have quite reliable distributions within the 95% confidence intervals for all months. A similar level of reliability was found in other catchments too. We also compare box plots of streamflow forecasts from the dynamic approach against those from the BJP (Figure 36 and Figure 37). These box-whisker plots show that variance of forecast distributions from the dynamic approach are comparable to those from the BJP and a visual comparison indicates that inter-quartile range is marginally lower than the range from BJP. Prior to bias correction, the dynamic model produced too emphatic forecasts due mainly to emphatic forecasts from POAMA1.5, which resulted in lower reliability. However, after the bias correction, streamflow outcomes from dynamic models were upgraded as both accurate and reliable forecasts.
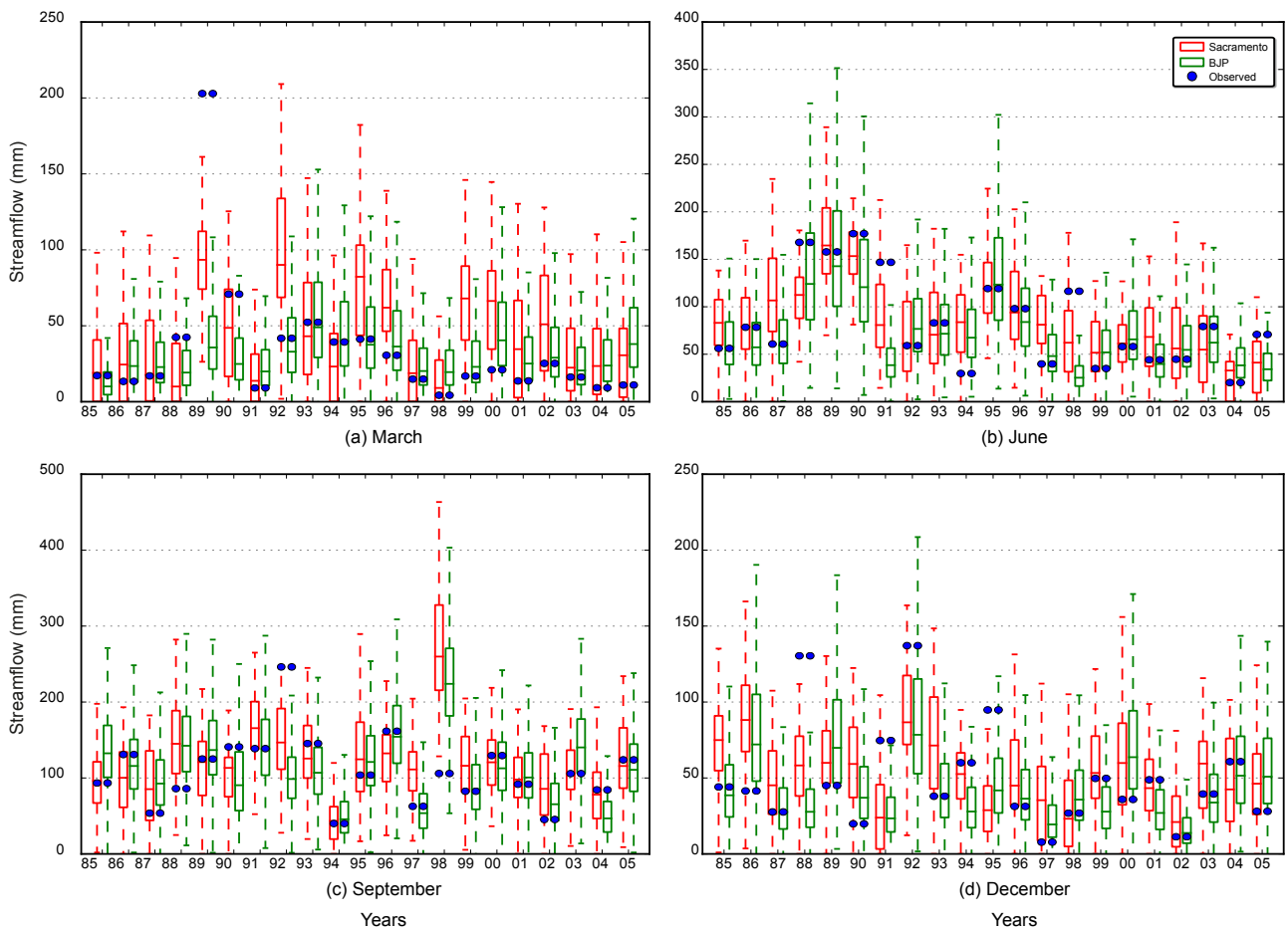
Figure 36: Seasonal streamflow forecasts from Sacramento with bias correction and BJP for Gingera
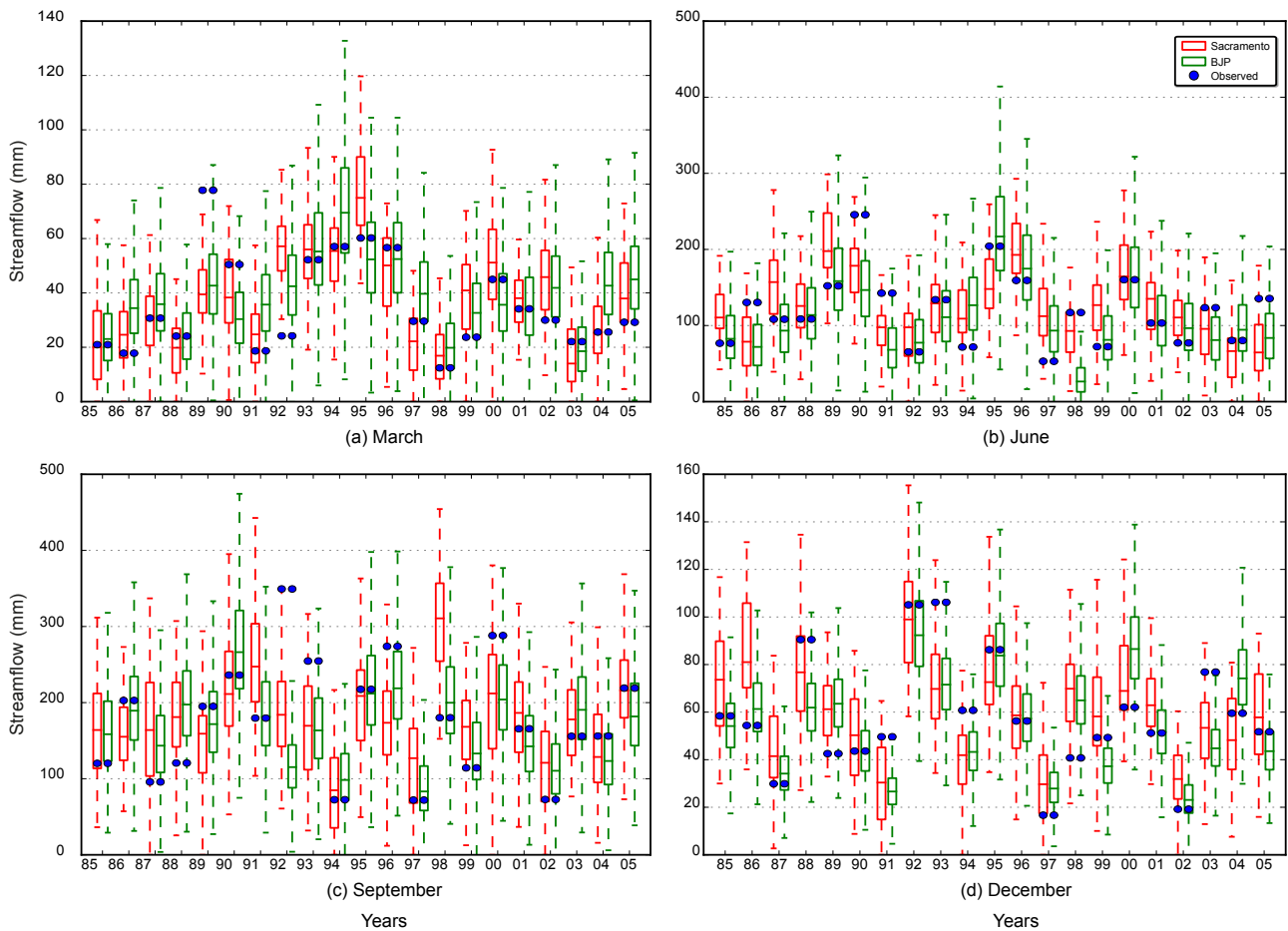
Figure 37: Seasonal streamflow forecasts from Sacramento with bias correction and BJP for Biggara

Considering the current scheme where Sacramento generated 55 samples of streamflow forecasts for each update (Figure 1) while BJP generated 5,000 samples, any statistics from such a small number of samples are relatively less certain than those derived from the BJP. Therefore, additional effort is required to obtain samples from a dynamic model. This will be achieved by: (a) enhancing the number of hydrologic model ensembles through the use of MCMC-based parameter samples using BATEA with predictive uncertainty (section 7), and (b) the use of 30 member ensemble rainfall forecasts from POAMA2.4a–c which are currently available from CAWCR. We believe that these two steps will help to generate more stable and potentially better estimates for both the accuracy and reliability of the dynamic modelling approach.

Figure 38 shows the hit/miss ratios of two percentile ranges of streamflow forecasts. Hit/miss ratios indicate how many times observed data lie within the inter-quartile range over the observed record length. For most catchments, the dynamic modelling approach provides slightly higher hit ratios than BJP in terms of both 25%–75% and 10%–90% ranges.

It is noteworthy that the skill scores in the figures are not strictly comparable with each other between the two modelling approaches. First, the skill scores of the dynamic modelling approach were calculated using the past observed data directly, while those of BJP were calculated using a statistically re-sampled distribution derived from the past observed data (Wang, Robertson & Chiew 2009). Even though the re-sampled distribution has a very similar pattern to the distribution of the observed data, the re-sampled distribution can reduce the impact of extreme values in observed data, which can increase skill score values, particularly ones on the probabilistic domain, even if slightly altered. Second, the model calibration and validation procedures adopted to generate the skill scores were different between the two modelling approaches.

For a more strict comparison, it is required for both methods to follow a single well-defined forecast verification procedure, which is planned as future work. Third, the skill scores of the dynamic modelling approach were calculated from 55 forecast ensemble members, while BJP generated 5,000 ensemble members. Therefore, as the ensemble member number of the dynamic modelling approach increases in future, it is possible that the estimate of model accuracy can be altered. Despite all the limitations, it is believed that the figures still provide valid results to compare the two modelling approaches.

We believe that the dynamic and statistical approaches are complementary to each other, and that blending outcomes from the two approaches can be a potential solution to improve both accuracy and reliability of the Bureau's SSF service significantly in the future.
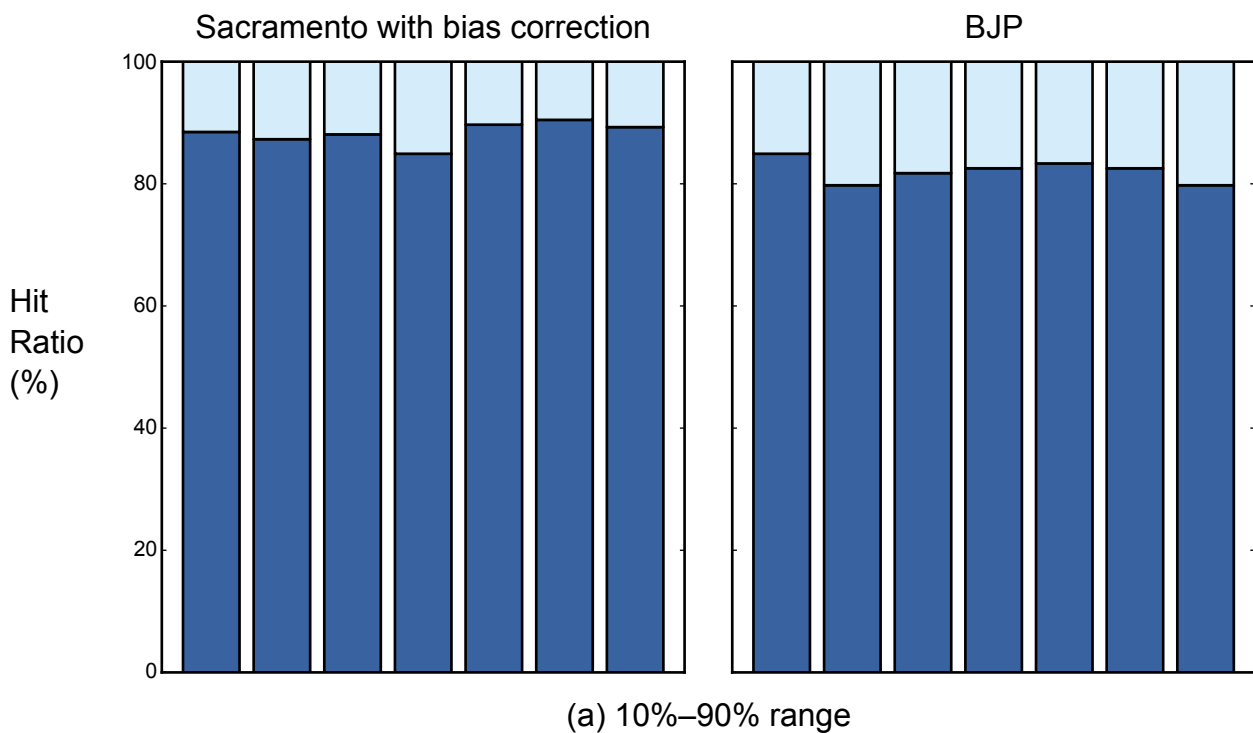
(a) 10%–90% range

Figure 38: Hit and miss ratios of seasonal streamflow forecasts from Sacramento and BJP

(b) 25%–75% range
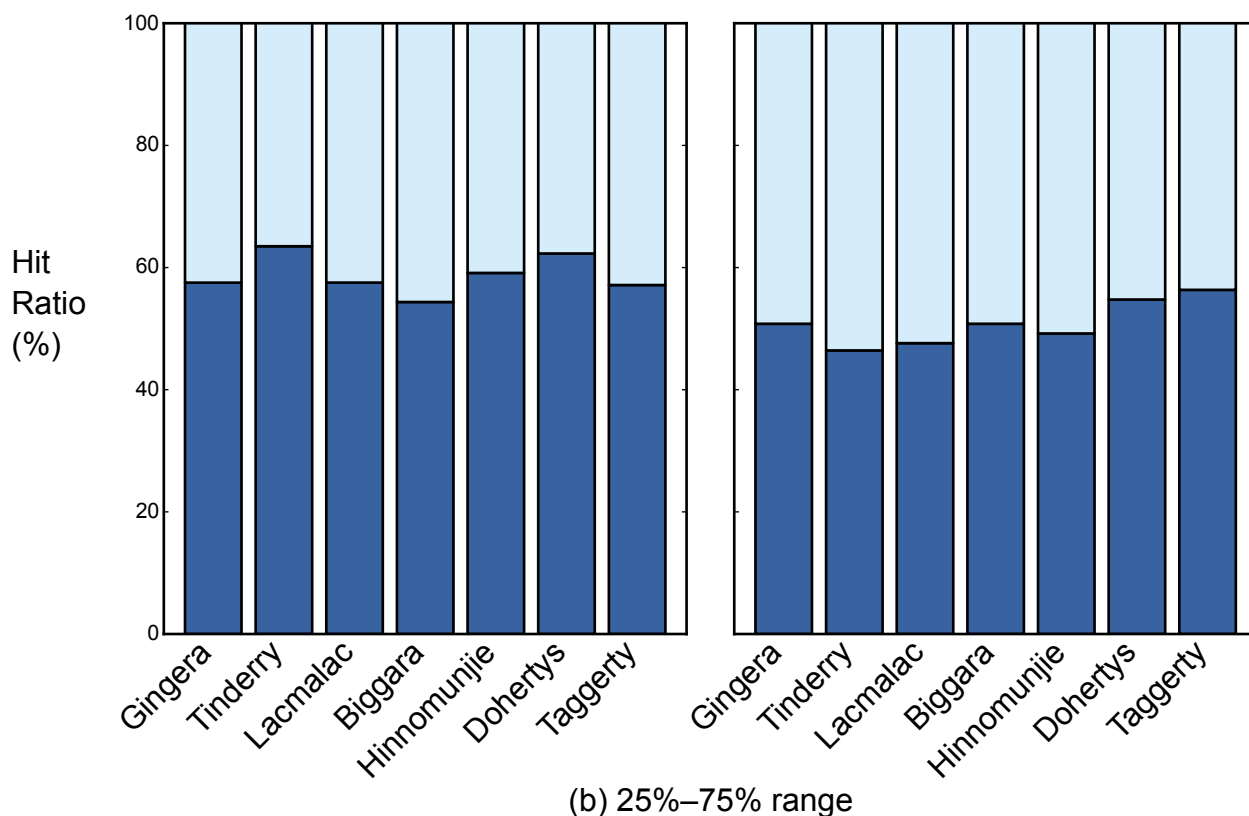
Figure 38: Hit and miss ratios of seasonal streamflow forecasts from Sacramento and BJP

6.6   Monthly versus seasonal forecast

So far, we have discussed seasonal (three-monthly) streamflow forecasts with an intention of finding additional value from the dynamic approach for the existing seasonal streamflow forecasting service. In our deliberations with stakeholders, the need to explore the possibility of a new monthly streamflow forecasting service updated at sub-monthly time steps was highlighted. In response to the need, we have also explored evaluation of the monthly streamflow forecasts updated at fortnightly time steps.

Noting that skills in POAMA forecasts are relatively high at one-month lead time in comparison to three-month lead time, the evaluation of monthly streamflow forecasts is an attractive option. Skill scores and NSE of monthly and seasonal (three-monthly) streamflow forecasts are compared in Figure 39. During February–March–April, the accuracy of monthly streamflow forecasts were not as good as that of seasonal streamflow forecasts, but during August–September–October–November monthly streamflow forecasts showed noticeable improvement in accuracy than the seasonal forecasts in most catchments. Monthly forecasts also showed slightly higher hit ratios than seasonal forecasts (Figure 40).
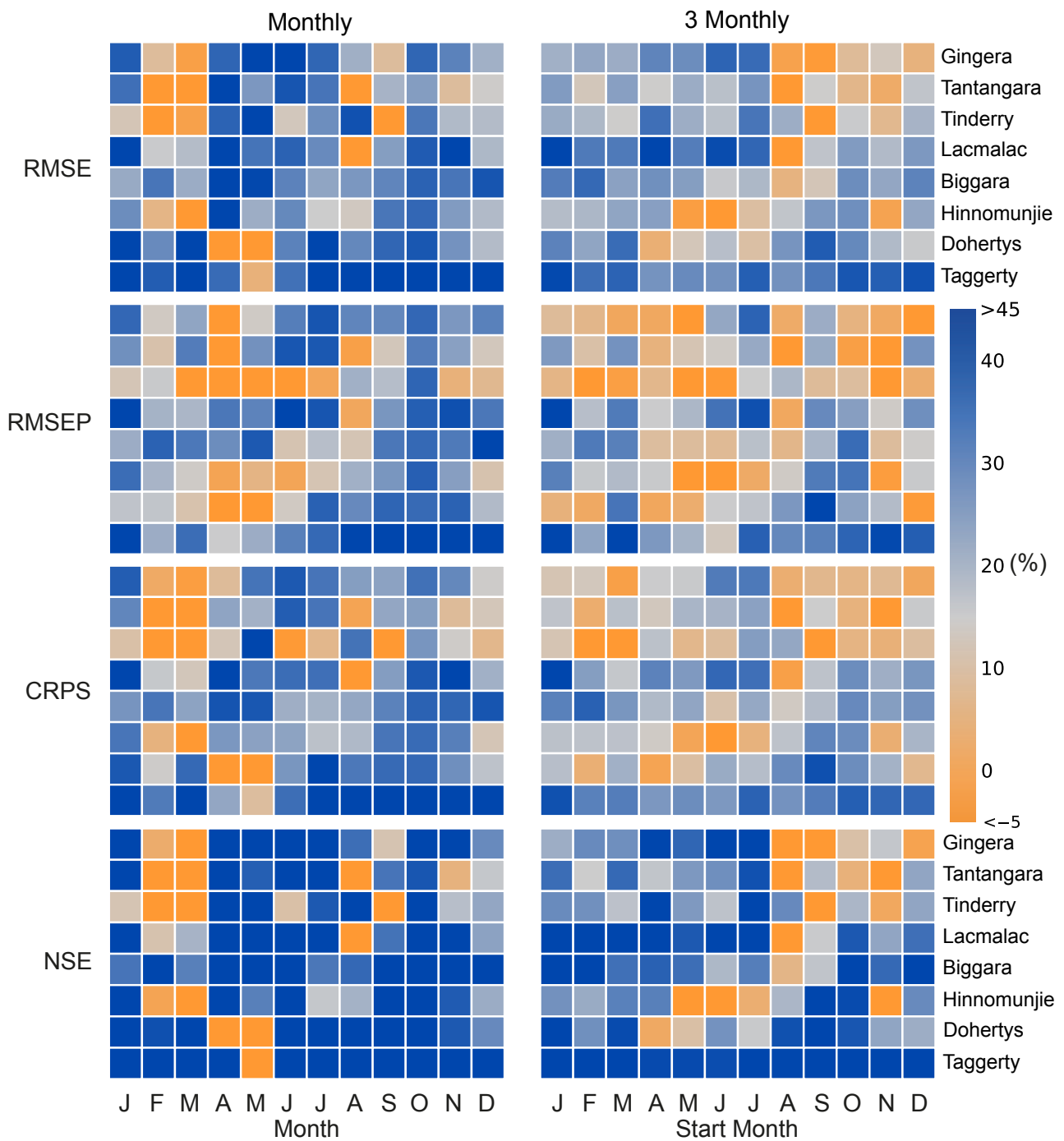
Figure 39: Skill scores, NSE and reliability of monthly and three-monthly streamflow forecasts. The forecasts were generated using Sacramento calibrated by the general parameterisation scheme with downscaled POAMA rainfall ensemble and then bias-corrected using ARMA method

Monthly

3 Monthly

(a) 10%–90% range
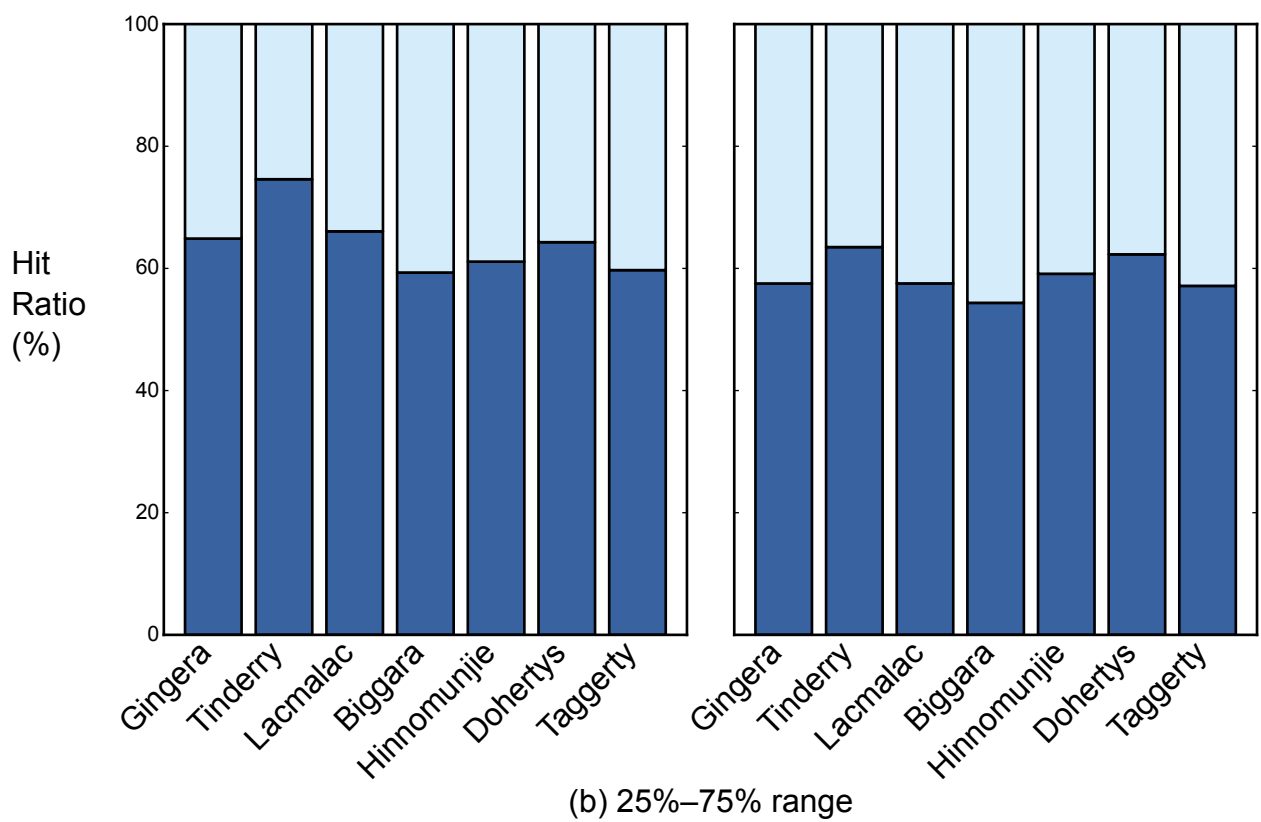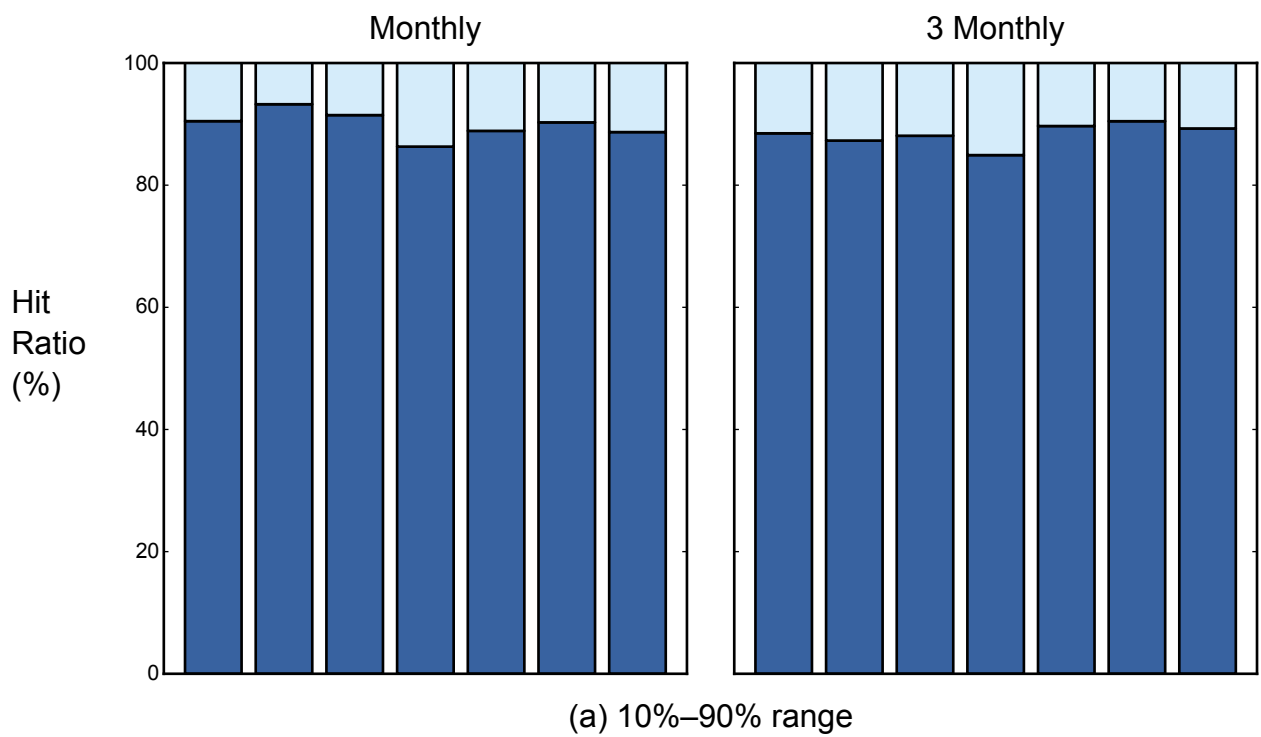
(b) 25%–75% range

Figure 40: Hit and miss ratios of monthly and three-monthly streamflow forecasts for 10–90% and 25–75% ranges

The main reasons for improved monthly streamflow forecasts are: (a) higher skills of POAMA at one-month lead time in comparison to three-month lead time during wet season, (b) stronger persistence in monthly streamflow data than seasonal streamflow data, and (c) relatively smaller drift in the hydrologic models, i.e. dynamic internal representation of the hydrologic model during the forecast period. In relative terms, we believe that enhanced forecast accuracy at monthly time steps is derived from persistence in streamflow data and, to a lesser degree, from more accurate rainfall forecasts. In addition, the stronger persistence enables the ARMA-based bias correction to more effectively remove biases in streamflow forecasts.

Predictive QQ plots or the PIT plots for Biggara (Figure 41) and Gingera (Figure 42) show that 'monthly' streamflow forecasts are as reliable as seasonal forecasts. A similar level of reliability was found in other catchments too.

The skill scores and reliability diagrams indicate that the dynamic approach can deliver accurate and reliable forecasts for monthly streamflow as well as seasonal streamflow. Considering the value of monthly streamflow forecasts to the Bureau's stakeholders, particularly those working on water supply operations, the possibility for the new service deserves more attention in the future. Further, we believe that continued development of POAMA, along with improved hydrologic modelling through BATEA, will enhance accuracy and reliability of the dynamic seasonal streamflow forecasts.
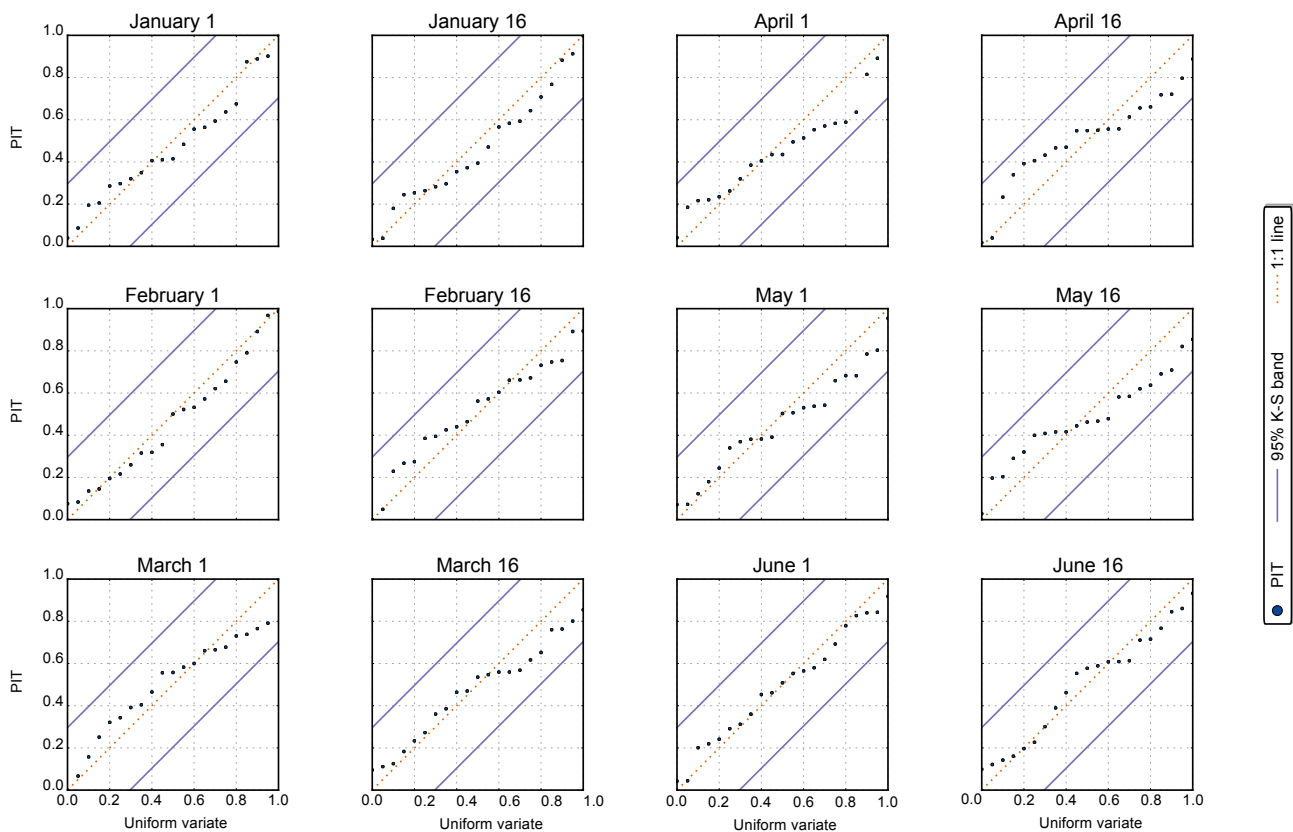
Figure 41: Predictive QQ plot of PITs from monthly streamflow forecasts for Biggara
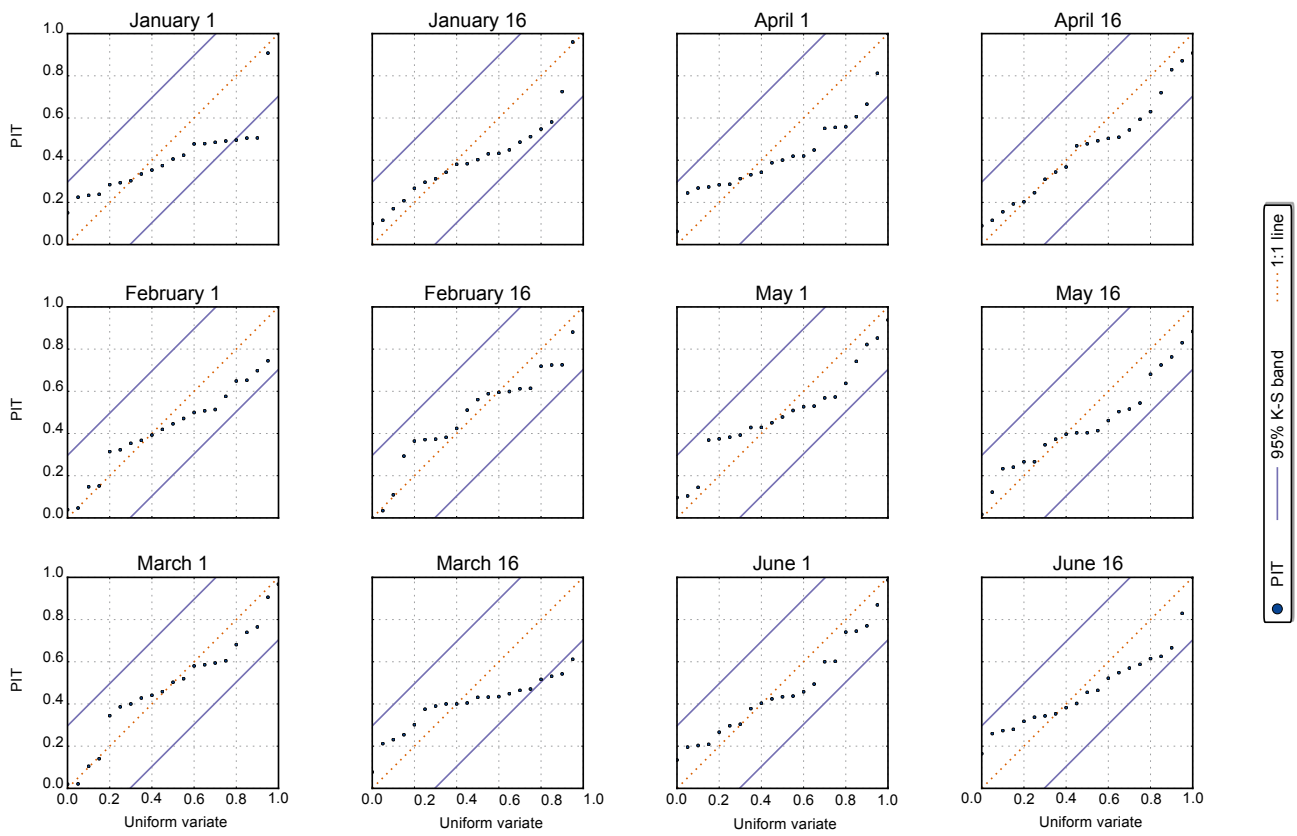
Figure 42: Predictive QQ plot of PITs from monthly streamflow forecasts for Gingera

7. Predictive uncertainty in seasonal forecasting and hydrologic modelling using BATEA

The Bureau has a keen interest in rigorous treatment of predictive uncertainty in hydrologic modelling, particularly in the context of seasonal forecasting. One such state-of-the-art technology available for estimation of predictive uncertainty is called the 'Bayesian Total Error Analysis (BATEA)' framework. The BATEA technology was developed by researchers from the University of Newcastle, Australia, and the University of Adelaide, Australia. The science underpinning BATEA has been extensively reviewed and published in leading hydrology and water resources journals (for further details see Kavetski, Kuczera & Franks 2006a; Kavetski, Kuczera & Franks 2006b; Kuczera et al. 2010a; Kuczera et al. 2010b; Renard et al. 2010; Thyer et al. 2009). It explicitly formulates the statistical models of errors in the observed input/output data and the model structure and exploits modern Bayesian techniques to process this information into probabilistic parameter estimates and model predictions. BATEA overcomes the known limitations of traditional methods such as Kalman filters and other such alternatives.

This section summarises major outcomes from the BATEA pilot project between the University of Newcastle (UoN), the University of Adelaide (UoA) and the Australian Bureau of Meteorology (the Bureau). A more detailed technical report on the BATEA project is currently under preparation by the research team. The summary is divided into two sections. The first section reports on the development of the BATEA software which is, ultimately, the principal deliverable to the Bureau. The second section describes the application of the evolving BATEA software to the selected Bureau's target experimental catchments.

7.1   BATEA software

The BATEA software implements the methods of the Bayesian Total Error Analysis calibration and prediction framework. During Stage 1 of the Bureau project, several major upgrades were carried out, including:

1. BATEA was linked to the Bureau's TIME-compliant rainfall-runoff models, SIMHYD and SACRAMENTO. In addition, the French model, GR4J was made available to the Bureau. It is noted that the numerical implementation used in SIMHYD to solve the water balance of the conceptual stores may not be robust, which results in artefacts and poor performance. These are related to the way water balance fluxes, especially those containing thresholds, are evaluated and propagated in time. It is beyond the scope of this project to address these issues, which appear common in the current generation of conceptual hydrological models [e.g. see reviews in Moore and Clarke (1981) and Kavetski and Clark (2010)]. Given that these numerical artefacts have the potential to degrade predictive performance and parameter estimation, it is recommended that the Bureau consider further investigations in the future to remove/reduce these numerical artefacts. Assistance with this can be provided if necessary by the BATEA team, subject to separate discussions and agreements.

2. A number of core calibration capabilities were completed and tested:

   a. BATEA offers a range of options for specifying error models describing input, output and model uncertainty. These error models define the objective functions used in the calibration. Operational level options include error models allowing for heteroscedasticity in the residual errors (the 'WLS' option). In general, a heteroscedastic error model refers to an error model that has a non-constant variance. For example, the simplest form is where the error variance increases linearly with the runoff. More advanced research level options already implemented include the use of general Bayesian hierarchical approaches to describe uncertainty in inputs (e.g. using rainfall error multipliers) and cater for time-varying stochastic rainfall-runoff model parameters. BATEA uses a combination of 'transformation' and 'probability' libraries to define the error models. Of importance here is that all libraries are scalable: new distributions and transformations can be added to this library as the need arises. With suitable training, this can be carried out by BATEA software users.

   b. The standard least squares (SLS) objective function used in the Bureau's Dynamic Modelling Systems (DMS) corresponds to a particular error model in BATEA (the 'SLS' option). Therefore, BATEA supports, as a special case, the objective function used for optimisation in the current version of DMS. However, BATEA provides more comprehensive tools for its analysis (e.g. optimisation, MCMC sampling, etc.).

   c. BATEA estimates parameter and predictive uncertainty in a statistically rigorous way, using Markov Chain Monte Carlo (MCMC) methods. It is noted that the multi-start optimisation approach used in DMS does not, in general, produce statistically rigorous estimates of parameter uncertainty. The DMS-derived parameter uncertainty estimates represent a measure of the lack of convergence of the optimisation algorithm and can fail for fairly simple test problems. In contrast, MCMC methods analyse the shape of the objective function using sampling techniques based on formal statistical theory and produce statistically rigorous estimates of parameter uncertainty.

   d. Several visualisation options are available within BATEA, including real-time visualisation of model response to changes in inputs and parameters. This can be used when drilling down and diagnosing problems in model performance. The visualisation can be turned off when carrying out batch runs from the command-line environment.

3. A major effort was invested in developing a comprehensive suite of post-processing diagnostics, named BAD (Bayesian Diagnostics). It is implemented using the R programming language. Its routines interrogate BATEA output files to produce a range of graphical outputs that report validation performance, evaluate the assumptions of the error models and visualise parameter uncertainty and the streamflow predictive uncertainty. These diagnostics have been partially automated within the BATEA software package as part of this Bureau project. In the future, further improvements, including more informative diagnostics, will be made available to the Bureau.

4. BATEA was developed to run in console mode using scripts to manage workflow. This will provide a pathway for workflow automation in the Bureau's systems (e.g. DMS). The scripts allow for commands and arguments. The script dictionary is scalable, i.e. adding new commands and arguments can be undertaken if/when necessary.

5. All outputs written to ASCII files that can be post-processed by the Bureau. Again, this simplifies the workflow automation within the Bureau's systems.

7.2 Application to the Bureau's experimental catchments

BATEA was applied to five catchments: Gingera, Tinderry, Biggara, Dohertys and Lacmalac using the SIMHYD, SACRAMENTO and GR4J models. In addition, an extensive analysis of rainfall sampling and runoff errors was conducted.
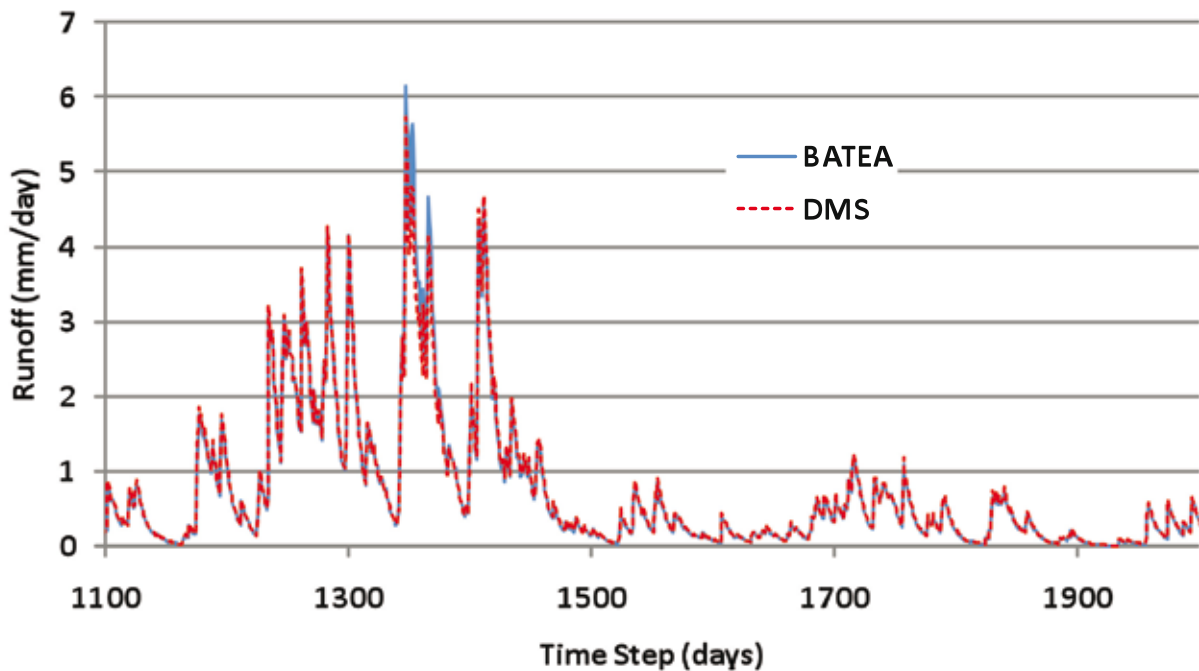
Figure 43: Comparison of simulated runoff for BATEA and DMS calibrations of SIMHYD model to Gingera catchment using standard least squares objective function

The main findings are listed below:

1. BATEA produces essentially the same results as DMS when the standard least squares (SLS) objective function is specified. This confirms that the software hook-up is operational (Figure 43).

2. The MCMC analysis of the parameter posterior distributions produced useful insights about the uncertainty in parameters (Figure 44):

a. For the SIMHYD model, insensitive parameters were identified which arise when the rainfall data do not force a process within the model. This is a particularly important issue when using the model in a predictive or forecast mode. For example, in some calibrations, the SIMHYD parameters associated with quickflow were insensitive because no quickflow was produced during the calibration. If the model was used with higher rainfall that activates the quickflow process, the prediction of quickflow is subject to potentially very large errors.

b. The MCMC analysis identifies highly correlated parameters, which are symptomatic of a poorly posed model (in particular, with non-identifiable parameters and model components).

c. The MCMC analysis provides a statistically rigorous assessment of parameter uncertainty, providing more meaningful estimates of parameter uncertainty than the multi-start optimisation strategy in DMS. The latter is actually characterising the difficulty of a particular algorithm in converging to a global optimum rather than providing a statistically rigorous parameter uncertainty. The MCMC approach overcomes these weaknesses of the multi-start model calibration approach used in DMS, and thus provides more reliable parameter estimates to be used in the Bureau's predictive applications. It should be noted that in most cases when using the SLS objective function, the parametric uncertainty is not the major contributor to the total predictive uncertainty. However, in case (a) preceding, parameters that were insensitive in calibration can become sensitive in validation, potentially generating very poor predictive performance. To an extent, this can be diagnosed using the MCMC analysis provided as part of the BATEA software.
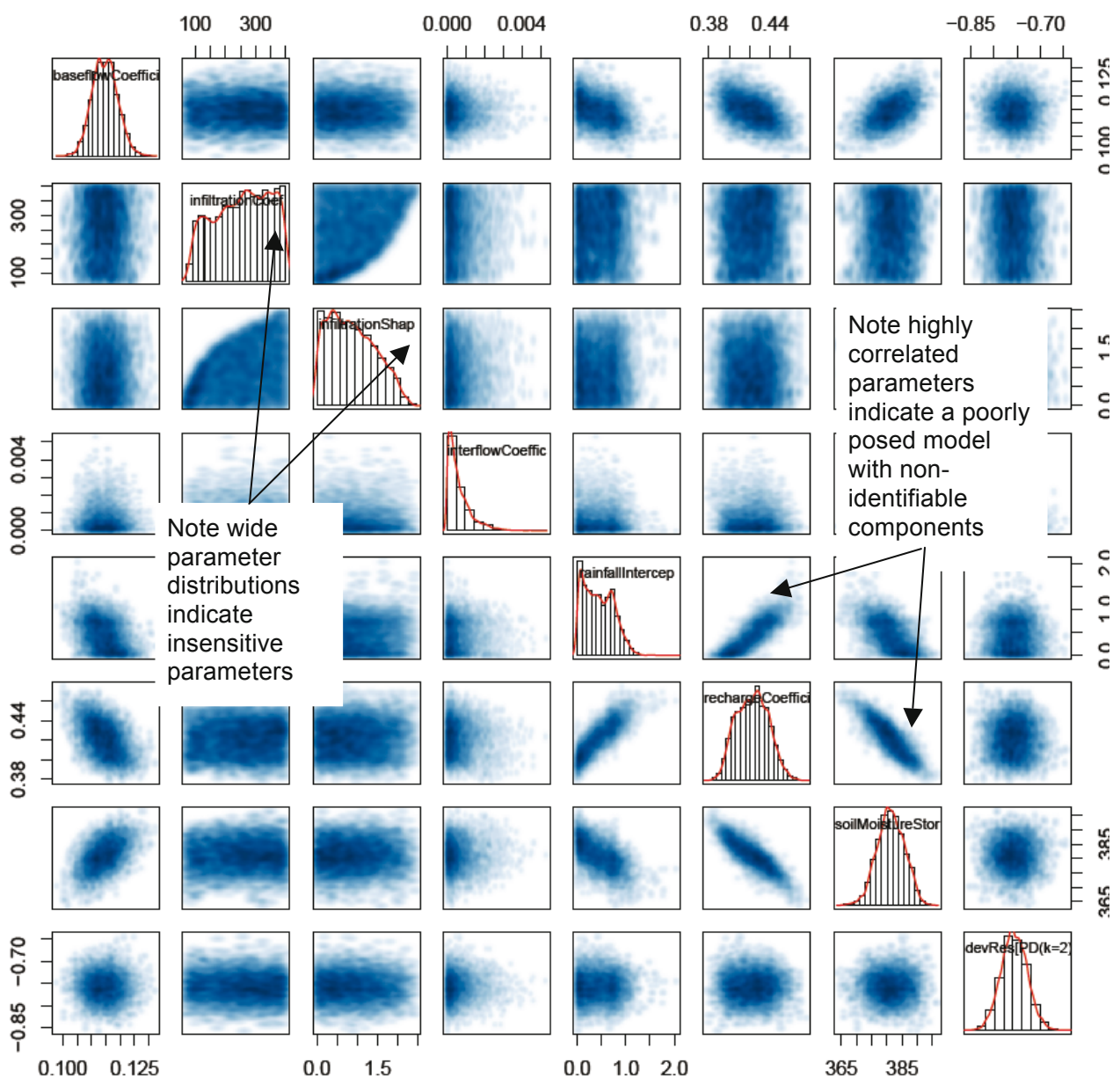
Figure 44: Parameter distributions for SIMHYD model calibrated to Gingera catchment using SLS. Plots on the diagonal provide the marginal parameter distributions. Off-diagonal plots provide the joint parameter distributions for each parameter pair (darker blue represents higher probability density)

3. Because a limited number of rain gauges are available, the error in catchment averaged daily rainfall can be quite considerable. A study of such errors was conducted using radar data to estimate the probability distribution of sampling errors in daily rainfall, conditioned on the magnitude of the daily rainfall. The results qualitatively agree with our analysis of a densely-gauged rainfall network in an experimental catchment in New Zealand and with some previous published findings (Figure 45). This provides additional confidence in these results.

Note the similar trend that the standard deviation of the rainfall error multipliers converges to a constant as observed rainfall increases. For Gingera, the standard deviation is higher because the number of rain gauges is lower than Tinderry and Lacmalac
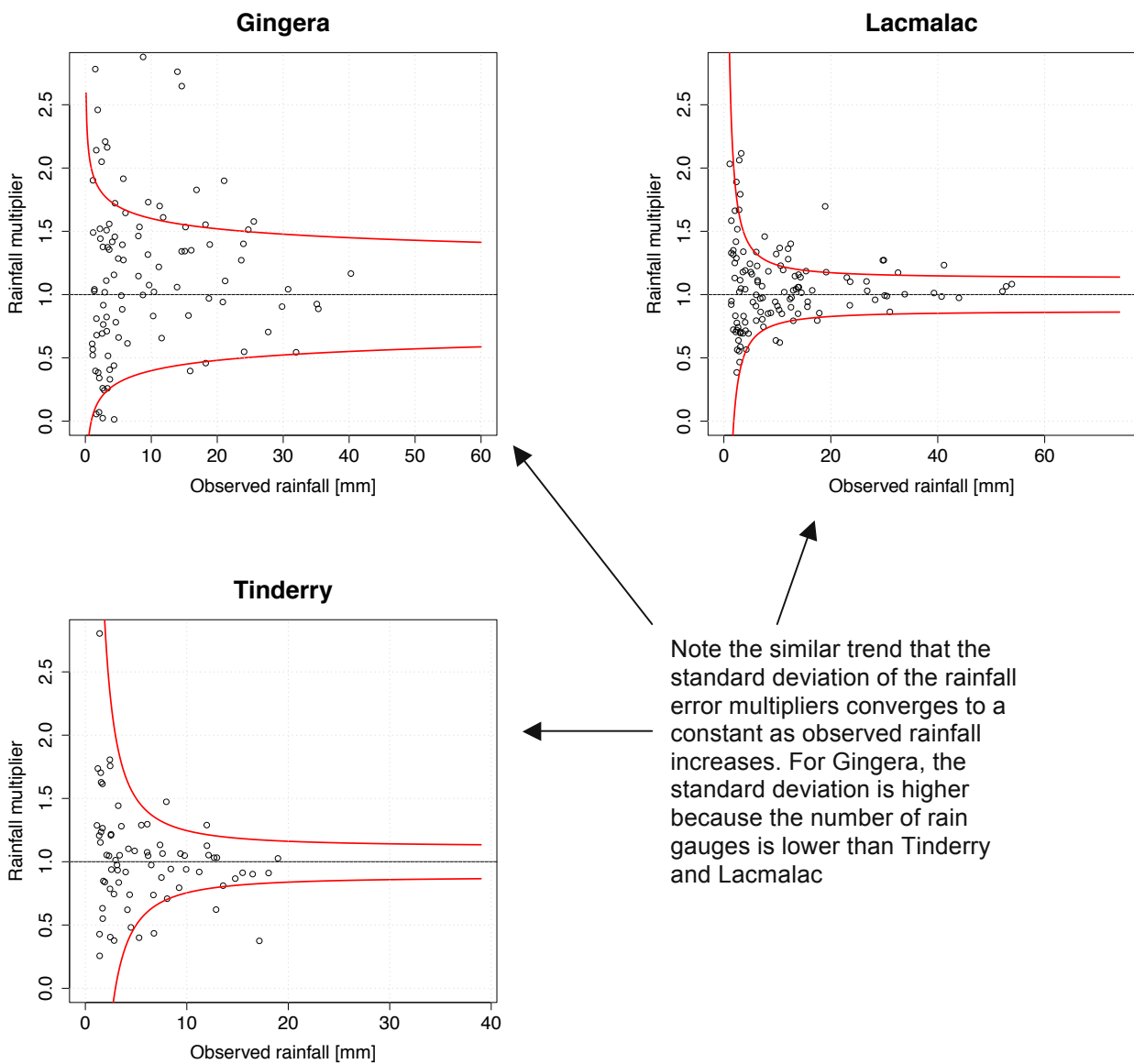
Figure 45: Multiplicative rainfall errors (f) estimated from radar data plotted as function of observed rainfall for Gingera, Lacmalac and Tinderry. The multiplicative rainfall error is defined as the ratio of the true areal rainfall over the observed rainfall. The red curves represent +/– the standard deviation of f which is estimated conditional on the observed rainfall

4. Analysis of runoff rating curve errors identified heteroscedastic error models where the standard deviation of the errors grows with the magnitude of runoff (see example in Figure 46). For the Bureau catchments, the standard deviation increased in the range 3–8% of the observed runoff. This is lower than the typically assumed 10% standard error and suggests the Bureau catchments are well gauged. A new characterisation of runoff error based on systematic errors in rating curve parameters was developed (Thyer et al. 2011). It will be incorporated into a future release of the BATEA software subsequent to this project.
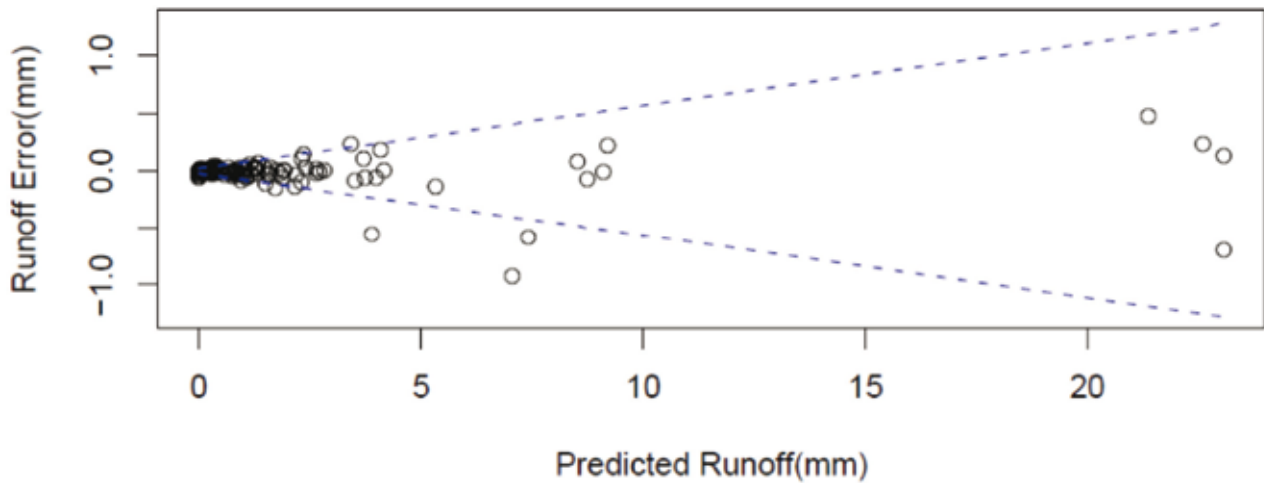
Figure 46: Runoff errors for Gingera catchment based on rating curve analysis. Dots represent the observed errors, based on difference between runoff gaugings and rating curve, while the blue dotted line represents the 95% probability limits of the fitted heteroscedastic runoff error model

5. The two strategic objectives for using the BATEA approach to calibration and prediction are to:

   i. Provide statistically reliable estimates of the predictive uncertainty. From the Bureau's perspective, this is important because predictions with reliable probability limits are needed for credible operational forecasting.

   ii. Decomposing the sources of the predictive uncertainty. If the contributions of the different sources of predictive uncertainty are known, this enables strategic investments to be made to reduce the predictive uncertainty.

Progress towards these objectives is as follows:

   a. In general, the use of heteroscedastic residual error model ('weighted least squares': WLS) in model calibration improves the reliability of the predictive uncertainty, relative to the SLS error model used in DMS. The degree of the improvement varies depending on catchment and hydrological model. Indeed, detailed analysis of high and flow streamflow ranges reveals some remaining deficiencies for low and high flows. This is likely due to the simple functional form (linear) used in the heteroscedastic residual error model. Further refinement and more reliable predictive uncertainty appear possible with the development of alternative functional forms (e.g. nonlinear) of the heteroscedastic residual error model and by accounting for runoff and rainfall error characteristics. This was only partially carried out in this project and is recommended for future testing and research.

   b. Previous research found that decomposing the uncertainty leads to significant improvement in the reliability of the predictive uncertainty (Renard et al. 2011). Technologically, this requires the specification of statistically reliable error models for the rainfall and runoff data. As part of this project, data analysis was carried out to estimate such models (see points 3 and 4 above). Investigations with incorporating runoff errors revealed similar predictive uncertainty to WLS. However, supplying the runoff error model can provide a lower bound on predictive errors and hence constitute an additional check of the model calibration. Although, for well gauged catchments the differences between the two versions of WLS can be minor, incorporating runoff error is a key step towards a decomposition of the sources of uncertainty. Investigations with using radar data to obtain independent estimates of catchment average rainfall error and using these estimates as precise priors for the rainfall errors, suggested that more research is needed for consistent improvements to be obtained within the context of large-scale automated analysis (whereas in previous published applications, it was possible to dedicate more resources to understand and model specific catchments and additional fieldwork data and insights were available). A major suspected reason is the omission of appropriate treatment of structural errors, which represents a major research issue in hydrological modelling. Though a number of potential strategies have been proposed in published work (including by the BATEA team), there is still significant research required before a suitable strategy can be recommended for operational use.

7.3 Summary and recommendation from the BATEA project

1. The BATEA software provides a range of parameter calibration and uncertainty analysis that are of considerable utility to the Bureau's hydrological modelling systems. These include MCMC parameter uncertainty estimation which can be used to diagnose model deficiencies, and a range of options for predictive uncertainty estimation. These comprise approaches that characterise the total predictive uncertainty, such as SLS and WLS and approaches which have the potential to decompose the sources of uncertainty. Implementing a technology that provides a wide range of options, including approaches that can be operationalised in the short-term, as well as research level approaches with significant potential in the longer term, is of clear benefit for the Bureau's forecasting systems.

2. From the range of options considered, the Weighted Least Squares (WLS) inference scheme appears best suited for operational use. It allows for variations in the total predictive error variance as a function of the streamflow and produces quite reliable predictions. It is also recommended to include runoff errors into the WLS scheme, as this is a key step towards the reliable decomposition of the sources of uncertainty. The appealing features of this approach are (i) its computational speed and (ii) methodological simplicity – in terms of both the underlying theory and the software settings. This means it is well suited as a first stage in the training of the use BATEA software for Bureau staff.

3. A full hierarchical treatment of input errors, and model structural errors, remains a research level endeavour. While the results are clearly encouraging in terms of improving the reliability of the predictive distribution of streamflow, and supporting the decomposition of predictive uncertainty into its contributing sources, it is a computationally more complex strategy and requires computational acceleration before it can be more routinely applied. It also necessitates the formulation of reliable error models describing uncertainties in the input data and appropriate treatment of structural errors. This can already be pursued for certain catchments, e.g. where dense gauges or radar coverage is available. However, it is not an automated procedure and further research and training is required to develop a strategy that can be applied in an operational framework across multiple catchments.

4. The BATEA software is compatible with TIME models. The current interfaces between BATEA and the TIME models are 'hard-coded' and are specific to SIMHYD and Sacramento. It is recommended that a more general interface be developed that allows choosing the time model to be connected to at runtime. This option is more scalable in terms of encompassing any future TIME model.

5. The GR4J model, available as a built-in option within the model library provided with the BATEA software, clearly outperforms SIMHYD for the Bureau catchments, and is generally competitive with Sacramento despite having only four calibrated parameters (as opposed to 20+ in Sacramento). Note that GR4J has been tested on thousands of catchments worldwide, and has shown good performance. Based on the result, it is planned to include GR4J as one of operational rainfall-runoff models for the SSF service.

6. It is envisaged that Stage 2 of the BATEA project will include training of Bureau staff and technical support to operationalise BATEA with WLS settings and runoff errors along with the MCMC analysis in a national rollout of the dynamic approach in the future. In addition, discussions were held regarding an ARC Linkage project which would allow tackling the research issues associated with the operational application of the 'full' BATEA including a hierarchical representation of input errors, model structural errors, recursive updating and other promising developments. These issues are crucial for the Bureau and will likely assist with enhancing reliability of the 'dynamic' seasonal streamflow forecasts.

7. Projects are developing between the BATEA team and industry partners such as CSIRO and DPI (Victoria). There is some synergy in these projects that can be exploited in terms of method and software development. Adoption of the BATEA software will provide the Bureau with a gateway to improved calibration and prediction algorithms, as the BATEA software engines undergo further developments. Importantly, at all times a range of options of different degrees of computational complexity and cost are available (of which WLS inference including runoff error characterisation and MCMC analysis is the currently recommended option for broader scale adoption).

8. Conclusion

A thorough evaluation of the monthly and three-monthly streamflow forecasting capability using the dynamic approach on eight target water supply catchments in the southern Murray–Darling Basin was completed by the Bureau and its research partners. As far as we are aware, this is the first comprehensive study in Australia that integrates the use of leading technologies through the application of POAMA1.5 rainfall forecasts, improved methods of seasonal downscaling and fairly well advanced techniques of hydrologic modelling and predictive uncertainty.

Three conceptual rainfall-runoff models used in this evaluation are: Sacramento, SIMHYD and SMAR. These models were calibrated and validated using a split sampling approach in ensemble mode. Except Tinderry and Hinnomunjie, these models showed good performance for both calibration (1976–1996) and validation (1998–2008) periods. The NSE from SIMHYD varied in the range 0.52 to 0.92 and 0.61 to 0.91 respectively, and that of Sacramento varied in the range 0.71 to 0.95 and 0.66 to 0.94.

Snowmelt is a significant component in the water budget of Hinnomunjie and Tantangara catchments. None of the models used includes accounting for snowmelt processes and therefore late winter and early spring runoff in these catchments was often under predicted. Hinnomunjie witnessed poor performance with all three models compared to other perennial catchments. For Tinderry, none of the models were able to reproduce streamflow patterns observed in the validation period, even though the models were calibrated to simulate streamflow patterns quite accurately during the calibration period. This is because Tinderry yielded extremely low streamflow during the recent drought, which coincided with the validation period used in the study.

Forecasting in simulation mode was performed for 1985–2005 with a warm-up period of 1980–1984. Rainfall forecasts from two sources were used in simulations: (a) downscaled POAMA rainfall ensemble (ten-member members and their mean), and (b) historical rainfall ensemble, which is the past ten-year observed rainfall data (ten-member members and their mean). Forecast skills were assessed using three skill scores: RMSE, RMSEP and CRPS. NSE was also estimated for forecast medians.

POAMA 1.5 rainfall forecast simulation datasets were downscaled using the improved analogue downscaling method to derive catchment scale rainfall from regional scale POAMA outputs. The relationship between POAMA rainfall and AWAP observed rainfall turned out to be weak even at monthly aggregated levels. Rainfall during wet months was frequently underestimated, which considerably reduced the variation range of POAMA rainfall compared to observed data. The underestimation of rainfall during wet seasons was a major source of bias in forecast streamflow outcomes from the rainfall-runoff models.

Comparison between the general parameterisation scheme (single parameter set) and the conditional parameterisation scheme (monthly varying parameter sets) using SIMHYD showed marginal differences in skill scores between the two schemes. Because of its more parsimonious hydrologic model representation, one single parameter set from the general parameterisation scheme was used in the subsequent evaluation.

Simulated streamflow forecast distributions from the historical rainfall ensembles and SIMHYD model were widely spread. However, when rainfall forcing was replaced with downscaled POAMA rainfall ensembles, in most years streamflow forecasts were underestimated. Even though downscaled POAMA rainfall and, accordingly, streamflow forecasts under predict observed streamflow resulting in poor skill scores, these forecasts do get the timing of highs and lows correct. Therefore, bias in streamflow forecasts is systematic and it is possible to capture this through a posterior streamflow bias correction procedure.

Posterior streamflow bias correction was implemented through the ARMA model on biased streamflow forecasts from the Sacramento model driven with downscaled POAMA 1.5 rainfall. Results from implementation of the ARMA model on target catchments at monthly and three-monthly timescales showed significant improvement in accuracy and reliability of the forecasts. The result suggests a strong case for the use of an ARMA-based bias correction procedure in dynamic seasonal streamflow forecasting and that a catchment specific ARMA model fitted across all update time steps will offer significant improvements in the streamflow forecasts.

Bias-corrected Sacramento streamflow forecasts driven through downscaled POAMA showed forecast skills comparable or better than those from the statistical BJP model. During January–July, three-monthly streamflow forecasts from the dynamic approach seem to outperform those from the BJP. However, BJP forecasts are more skilful during August–December, particularly for Gingera and Tinderry, while skills are comparable in other catchments. Therefore, we believe that blended streamflow forecasts derived from a combination of statistical and dynamic approaches in future work are likely to provide better streamflow forecasts.

We compared monthly and three-monthly streamflow forecasting capability from the dynamic approach and found the dynamic modelling approach can deliver monthly streamflow forecasts as accurate and reliable as three-monthly streamflow forecasts.

As an additional deliverable, a modelling system called the Dynamic Modelling System (DMS) was developed, and its architecture is currently under transition to an operational system.

As a rigorous method for predictive uncertainty estimation, BATEA was applied to target catchments. Using the Bayesian-based Markov Chain Monte Carlo (MCMC) method, BATEA provided more realistic estimates of predictive uncertainty as well as diagnostic evidence for each error source including model structural deficiency and rating curve errors. The results indicated that the technology has a potential to improve not only seasonal streamflow forecast service but also many other services of the Bureau in terms of predictive uncertainty estimation.

Another key outcome of the BATEA strand of the project is that further research could be conducted on approaches that decompose the distinct sources of uncertainty affecting the forecasting system. The ability to decompose the sources of uncertainty not only leads to more reliable predictions, but also provides strategic guidance on how to reduce predictive uncertainty. It is better suited for taking advantage of improvements in various components of the flow forecasting system, when compared to approaches that focus only on characterising the aggregate uncertainty.

9. Recommendation for operational service and further research

The Experimental evaluation of the Dynamic Seasonal Streamflow Forecasting approach has now been completed successfully. On the basis of detailed findings from the work reported in sections 6 to 8 of this report, we recommend the following future directions for major enhancements to the current Seasonal Streamflow Forecasting (SSF) service:

1. A national rollout of the SSF service across Australia beyond the current 21 target water supply catchments within the Murray–Darling Basin.

2. The current ARMA-based bias correction turned out an effective solution to improve both the accuracy and reliability of streamflow forecasts from conceptual rainfall-runoff models. Further research and evaluation will likely be required for those catchments where the posterior bias correction strategy does not produce accurate and reliable forecasts.

3. Identification of high quality streamflow reference stations suitable for the enhanced SSF service. We note that this work has already been agreed to in principle by the Bureau and that further activity planning is well advanced.

4. Further develop the communications and adoption strategy for major enhancements to the current service including a two-stage stakeholder engagement approach to decide on target locations for the national rollout.

5. Develop and implement the methodology to provide blended three-month lead time seasonal streamflow forecasts derived from both the statistical and the dynamic approaches. The methodology could be developed by the WIRADA research team and implemented by the Bureau staff.

6. A new national one-month lead time streamflow forecasting service based on products derived from the dynamic approach in addition to the proposed blended three-monthly streamflow forecasts. Note that the work has shown the possibility of accurate and reliable monthly forecasts. Further, there is a demand for a monthly streamflow forecasting service updated at frequent sub-monthly time intervals from operational water management agencies (e.g. MDBA) and the proposed service will help address this demand.

7. Further develop and operationalise the experimental Dynamic Modelling System (DMS) for use in the national rollout of the service, including additional development and linkages with the BATEA system for handling predictive uncertainty.

8. Use of model structures that include the capability to model snowmelt processes, particularly in high runoff generation areas in the southern Alpine region of the Murray–Darling basin. Use of the Snow17 model of the US National Weather Service is a possible option.

9. Further targeted research by WIRADA and CAWCR research teams for seeking more accurate and reliable rainfall forecasts. This could be achieved through a number of activities such as the inclusion of seasonal climate forecasts from multiple GCMs and further improvements in POAMA rainfall forecasts and its downscaling methods. There was also recent progress in applying BJP for the post-processing of POAMA rainfall forecasts.

10. The Weighted Least Squares (WLS) scheme implemented by the BATEA technology is recommended for evaluating the predictive uncertainty in an operational context. The advantage of using BATEA technology is that it provides not only ability to undertake model diagnosis of dynamic model calibration, but that it provides support for a wide range of options for predictive uncertainty analysis, including operational level, and research level techniques (see following recommendation). As such, it serves as an umbrella for a wide range of methods and the selection of methods from the available options is a decision for the Bureau in consultation with the BATEA team.

11. Further research is recommended on the hierarchical representation of input and model structural errors. This has strategic long-term benefits over a range of forecast lead times, both in terms of improving the statistical reliability of the dynamic seasonal streamflow forecasts, and reducing the uncertainty in forecasts through a better assimilation of the information content in the data.

Appendix A

Specifications for ARMA modelling to be used for error updating in forecasting work Experimental Dynamic Seasonal Forecasting.

Version 1 – Prepared by Narendra Kumar Tuteja and Sri Srikanthan (2 December 2010)

Version 2 – Prepared following comments from Sri Srikanthan and George Kuczera (8 December 2010)

Forecasting work in simulation mode used downscaled POAMA rainfall (11 members comprising ten POAMA ensembles and an ensemble mean) and five hydrological model ensembles each corresponding to Sacramento and SIMHYD models for the period 1980–2006. The two products that explored using the dynamic seasonal streamflow forecasting approach are:

- Monthly streamflow forecasts updated fortnightly ($qmon_{1980\text{-}01\text{-}01}$, $qmon_{1980\text{-}01\text{-}16}$, $qmon_{1980\text{-}02\text{-}01}$, $qmon_{1980\text{-}02\text{-}16}$ ......)

- Three-monthly streamflow forecasts updated monthly ($q3mon_{1980\text{-}01\text{-}01}$, $q3mon_{1980\text{-}02\text{-}01}$, $q3mon_{1980\text{-}03\text{-}01}$, $q3mon_{1980\text{-}04\text{-}01}$ ......).

If successful, these forecast products in simulation mode will be converted into products in forecast mode.

The four skill scores used to assess performance of the forecast streamflows are: SSRMSE, SSRMSEP, CPRS and seasonal NSE (with respect to respective median historical reference streamflows for both monthly and three-monthly forecast products). Further, reliability plots are developed for assessing skills of the forecast probability distribution (also referred to as PIT plots or QQ plots). Note that a common Forecast Verification tool (FCVF) was used to assess skills of the forecasts from dynamic and statistical systems.

Streamflow forecasts simulated from the best hydrologic model ensemble and the ensemble mean rainfall include bias in streamflow resulting from the following sources: inaccuracy in climate forcing (POAMA rainfall, downscaling and evaporative demand), inaccuracy in streamflow measurements and structural inadequacy in hydrologic model and parameter uncertainties. Since streamflow is a damped response, we expect persistence in the streamflow bias time series. ARMA (p,q) model needs to be fitted to the following monthly $qmon$ and three-monthly forecast streamflow time series $q3mon$ to obtain bias corrected streamflow forecasts:

- forecast streamflow time series corresponding to ensemble mean rainfall and the best hydrologic model, say $q_t = \hat{Q}\left(p_e = 11\right)$

- forecast streamflow time series corresponding to the median of the 55 hydrologic model ensembles, say $q_t = \hat{Q}median\left(55\,\text{ensembles}\right)$.

Note that to avoid data overlap for error updating, two separate ARMA models are required for monthly forecasts updated fortnightly and three separate ARMA models are required for three-monthly forecasts updated monthly.

The following notation is used in describing the ARMA (p,q) model fitting:

$p =$ order of autoregressive AR component of the ARMA model (eq 1)

$q =$ order of moving average MA component of the ARMA model (eq 2)

$q_t =$ forecast streamflow time series (monthly or 3-monthly time series corresponding to $\hat{Q}(p_e = 11)$ or $\hat{Q}median(55\ ensembles)$ (eq 3)

$x_t = q_{obs} - q_t =$ forecast streamflow bias time series (monthly or three-monthly time series corresponding to $\hat{Q}(p_e = 11)$ or $\hat{Q}median(55\ ensembles)$ (eq 4)

$\mu_T = \dfrac{\sum\limits_{i=1}^{n} x_{t,T}}{n} =$ periodic mean flow of the respective

streamflow bias time series x_t ($xmon$ for T = Jan-1,Jan-16,Feb-1,Feb-16,…..Dec-16 or $x3mon$ for T = Jan-1,Feb-1,…..Dec-1) (eq 5)

$\sigma_T = \sqrt{\dfrac{\sum\limits_{j=1}^{n} (x_{t,T} - x_T)}{n-1}} =$ periodic standard deviation of

the respective streamflow bias time series x_t ($xmon$ or $x3mon$; ideally these need to be calculated from Fourier Transforms but for now we will use those obtained from raw data) (eq 6)

$y_t = \dfrac{x_t - \mu_T}{\sigma_T} =$ standardised streamflow bias time series N (0,1) (eq 7)

Form of the ARMA (p,q) model:

$$\left(1 - \sum_{i=1}^{p} \phi_i B^i\right) y_t = \left(1 + \sum_{j=1}^{q} \theta_j B^j\right)\varepsilon_t \text{ (eq 8)}$$

$$\Phi(B) y_t = \Theta(B)\varepsilon_t \text{ (eq 9)}$$

Where, $i = 1,2,....p =$ order of the AR component, $j = 1,2,....q =$ order of the MA component, $\varepsilon_t =$ standard normal variate N $(0, \sigma_\varepsilon^2)$, σ^2 residual variance i.e. remaining variance after bias correction, $B =$ backward shift operator, $\Phi(B) = 1 - \sum\limits_{i=1}^{p} \phi_i B^i$ and

$\Theta(B) = 1 + \sum\limits_{j=1}^{q} \theta_j B^j$

Steps involved in fitting the ARMA model are described below:

1. Calculate streamflow bias time series x_t from eq4

2. Calculate periodic mean of streamflow bias time series μ_T from eq5

3. Calculate periodic standard deviation time series σ_T from eq6

4. Calculate standardised streamflow bias time series y_t from eq7

5. Calculate statistics of the complete forecast time series: mean, standard deviation, skewness (normality) and ACF of the q_t, x_t & y_t

6. Develop ACF of the standardised streamflow bias time series y_t up to two years lag (48 for monthly time series updated fortnightly and 24 for three-monthly time series updated monthly)

7. Develop PACF (Partial Autocorrelation Function) of the standardised streamflow bias time series y_t up to order k = 10

8. Calibrate the following ARMA (p,q) models (0,1), (0,2), (1,0), (2,0), (3,0), 4,0), (1,1), (2,1), (3,1), 4,1), (1,2), (2,2), (3,2) and (4,2) using *Akaike Information Criteria* (AIC) as the objective function (Box & Jenkins 1976; Hyndman 2011). Using mean squared error MSE, estimate standard deviation of the residual error $\hat{\sigma}_\varepsilon = \sqrt{MSE}$ (eq 10)

9. Identify and adopt a suitable model form amongst the candidate models based on outcomes of the calibrated models in step 8.

10. Simulate streamflow bias $\hat{y}_t$ from the adopted ARMA (p,q) model using (note that ε_t is the standard normal variate N $(0, \sigma_\varepsilon^2)$ $\left(1 - \phi_i B^i\right) y_t = \left(1 + \theta_j B^j\right)\varepsilon_t$ (eq 11)

11. Simulate $\hat{x}_t = \mu_T + \sigma_T \hat{y}_t$ (eq 12)

12. Simulate updated forecast streamflow $\hat{q}_t = q_t + \hat{x}_t$ (eq13)

13. Do forecasting for the period 1980–2006 (see table below)

14. Calculate skill scores SSRMSE and NSE_seasonal using $\hat{q}_t$.

Table 15 Example of the forecasting process for an ARMA (1,1) model

Date	y_t	$\hat{y}_t$	$\hat{\varepsilon}_t$
1985-01-01	y_1	$\hat{y}_1 = 0$	$\hat{\varepsilon}_1 = y_1 - \hat{y}_1$
1985-02-01	y_2	$\hat{y}_2 = \phi_1 y_1 + \varepsilon_2 + \theta_1 \hat{\varepsilon}_1$	$\hat{\varepsilon}_2 = y_2 - \hat{y}_2$
1985-03-01	y_3	$\hat{y}_3 = \phi_1 y_2 + \varepsilon_3 + \theta_1 \hat{\varepsilon}_2$	$\hat{\varepsilon}_3 = y_3 - \hat{y}_3$
....			
....			
2005-12-01	y_n	$\hat{y}_n = \phi_1 y_{n-1} + \varepsilon_n + \theta_1 \hat{\varepsilon}_{n-1}$	$\hat{\varepsilon}_n = y_n - \hat{y}_n$

Acronyms

ACF	Autocorrelation function
ARMA	Autoregressive moving-average
AWAP	Australia Water Availability Project
AWRIS	Australian Water Resources Information System
BATEA	Bayesian total error analysis
BJP	Bayesian joint probability
CAWCR	Centre for Australian Weather and Climate Research
CRPS	Continuous ranked probability score
CSIRO	Commonwealth Scientific and Industrial Research Organisation
CWIPIT	Climate and Water Information Program Information Technology
DMS	Dynamic Modelling System
DPI	Department of Primary Industries
EHP	Extended hydrological prediction
GCM	Global Climate Model
MCMC	Markov Chain Monte Carlo
MDBA	Murray–Darling Basin Authority
MSE	Mean square error
NSE	Nash Sutcliffe efficiency
PACF	Partial autocorrelation function
PET	Potential evapotranspiration
PIT	Probability integral transform
POAMA	Predictive Ocean Atmosphere Model for Australia
QQ	Quantile–Quantile
RMSE	Root mean–squared errors
RMSEP	Root mean-squared errors in probability
SDM	Statistical Downscaling Model
SEACI	South Eastern Australian Climate Initiative
SOI	Southern Oscillation Index
SSF	Seasonal streamflow forecast
SST	Sea surface temperature
SEWPaC	Department of Sustainability, Environment, Water, Population and Communities
SLS	Standard least square
UMC	Upper Murray Catchment
WAFARi	Water Availability Forecast of Australian Rivers
WLS	Weighted least square
WIRADA	Water Information Research and Development Alliance

References

Alves, O, Wang, G, Zhong, A, Smith, N, Tseitkin, F, Warren, G, Schiller, A, Godfrey, S & Meyers, G 2003, POAMA: Bureau of Meteorology operational coupled model seasonal forecast system, *National Drought Forum*, Brisbane.

Beven, KJ & Kirkby, MJ 1979, 'A physically based, variable contributing area model of basin hydrology', *Hydrological Sciences Bulletin*, vol. 24, pp. 43–69.

Box, GEP & Jenkins, GM 1976, *Time series analysis: Forecasting and control, revised edition*, Holden-Day, San Francisco.

Burnash, RJ 1985, 'The Sacramento watershed model, hydrologic modelling for flood hazards' *Hydrologic Modelling for Flood Hazards Planning and Management*, University of Colorado at Denver, California-Nevada River Forecast Center, National Weather Service, Sacramento.

Burnash, RJC, Ferral, RL, McGuire, RA & Joint Federal-State River Forecast Center 1973, *A generalized streamflow simulation system: conceptual modeling for digital computers*, U. S. Dept. of Commerce, National Weather Service.

Charles, A, Fernandez, E, Timbal, B & Hendon, H 2010, *Analogue downscaling of POAMA seasonal rainfall forecasts, Technical Report for review*, Climate and Water Division, Bureau of Meteorology, Australia.

Chiew, F & McMahon, T 1994, 'Application of the daily rainfall-runoff model MODHYDROLOG to 28 Australian catchments', *Journal of Hydrology*, vol. 153, pp. 383–416.

Chiew, FHS, Peel, MC & Western, AW 2002, 'Application and testing of the simple rainfall-runoff model SIMHYD', in VP Singh & DK Frevert (eds), *Mathematical Models of Small Watershed Hydrology and Applications*, Littleton, Colorado, Water Resources Publication.

Chiew, FHS, Vaze, J, Viney, NR, Perraud, JM, Teng, J & Post, DA 2008, 'Modelling runoff and climate change impact on runoff in 178 catchments in the Murray–Darling basin using Sacramento and SIMHYD rainfall-runoff models', in M Lambert, TM Daniell & M Leonard (eds), *Proceedings of Water Down Under 2008*, Adelaide.

Chiew, FHS, Zhou, SL & McMahon, TA 2003, 'Use of seasonal streamflow forecasts in water resources management', *Journal of Hydrology*, vol. 270, pp. 135–144.

Clewett, JF, Clarkson, NM, George, DA, Ooi, S, Owens, DT, Partridge, IJ & Simpson, GB 2003, *Rainman streamFlow (version 4.3): A comprehensive climate and streamflow analysis package on CD to assess seasonal forecasts and manage climatic risk*, Department of Primary Industries, Brisbane.

Crawford, NH & Linsley, RK 1966, *Digital simulation in hydrology Stanford watershed model 4, Tech Report*, 39th edn, Department of Civil Engineering, Stanford University.

Dawid, AP 1984, 'Present position and potential developments: Some personal views: Statistical theory: The prequential approach', *Journal of the Royal Statistical Society, Series A (General)*, vol. 147, pp. 278–292.

Day, GN 1985, 'Extended streamflow forecasting using NWSRFS', *Journal of Water Resources Planning and Management*, vol. 111, pp. 157–170.

De Gooijer, JG & Zerom, D 2000, 'Kernel-based multistep-ahead predictions of the US short-term interest rate', *Journal of Forecasting*, vol. 19, pp. 335–353.

Gneiting, T, Balabdaoui, F & Raftery, AE 2007, 'Probabilistic forecasts, calibration and sharpness', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 69, pp. 243–268.

Gómez, V 1998, *Automatic model identification in the presence of missing observations and outliers*, Ministerio de Economía y Hacienda, Dirección General de Análisis y Programación Presupuestaria, Madrid.

Gómez, V & Maravall, A 1997, *Programs TRAMO and SEATS: Instructions for the User*, Ministerio de Economia y Hacienda, Dirección General de Análisis y Programación Presupuestaria, Madrid.

Goodrich, RL 2000, 'The Forecast Pro methodology', *International Journal of Forecasting*, vol. 16, pp. 533–535.

Hammer, GL, Nicholls, N & Mitchell, CD 2000, *Applications of seasonal climate forecasting in agricultural and natural ecosystems: The Australian experience*, Kluwer Academic Publishers, The Netherlands.

Hannan, EJ & Rissanen, J 1982, 'Recursive estimation of mixed autoregressive-moving average order', *Biometrika*, vol. 69, pp. 81–94.

Hyndman, RJ 2011, *forecast: Forecasting functions for time series*, viewed 2 September 2011, http://cran.r-project.org/web/packages/forecast/index.html

Hyndman, RJ & Khandakar, Y 2008, 'Automatic Time Series Forecasting: The forecast Package for R', *Journal of Statistical Software*, vol. 27, pp. 1–22.

Kachroo, RK 1992, 'River flow forecasting, Part 5: Applications of a conceptual model', *Journal of Hydrology*, vol. 133, pp. 141–178.

Kachroo, RK & Liang, GC 1992, 'River flow forecasting, Part 2: Algebraic development of linear modelling techniques', *Journal of Hydrology*, vol. 133, pp. 17–40.

Kavetski, D & Clark, MP 2010, 'Ancient numerical daemons of conceptual hydrological modeling: 2, Impact of time stepping schemes on model analysis and prediction', *Water Resources Research*, vol. 46, W10511.

Kavetski, D, Kuczera, G & Franks, SW 2006a, 'Bayesian analysis of input uncertainty in hydrological modeling: 1, Theory', *Water Resources Research*, vol. 42, W03407.

Kavetski, D, Kuczera, G & Franks, SW 2006b, 'Bayesian analysis of input uncertainty in hydrological modeling: 2, Application', *Water Resources Research*, vol. 42, W03408.

Krause, P, Boyle, DP & Bäse, F 2005, 'Comparison of different efficiency criteria for hydrological model assessment', *Advances in Geosciences*, vol. 5, pp. 89–97.

Kuczera, G, Kavetski, D, Renard, B & Thyer, M 2010a, A limited-memory acceleration strategy for MCMC sampling in hierarchical Bayesian calibration of hydrological models. *Water Resources Research*, vol. 46, W07602.

Kuczera, G, Renard, B, Thyer, M & Kavetski, D 2010b, 'There are no hydrological monsters, just models and observations with large uncertainties!', *Hydrological Sciences Journal*, vol. 55, pp. 980–991.

Laio, F & Tamea, S 2007, 'Verification tools for probabilistic forecasts of continuous hydrological variables', *Hydrology and Earth System Sciences*, vol. 11, pp. 1267–1277.

Laugesen, R 2010, *Seasonal dynamic modelling system (DMS), v.0.8.2*, Water Division, Bureau of Meteorology, Australia.

Laugesen, R, Shin, D & Tuteja, NK 2009, *Pilot seasonal dynamic modelling system specification and proposed solution*, Water Forecasting Branch, Bureau of Meteorology, Australia.

Hendon, HH & Alves, O 2009, Assessment of improved POAMA forecasts for hydrologically relevant surface variables. *Final report for South-Eastern Australian Climate Initiative (SEACI), project 3.1.2P*, viewed 2 September, http://www.seaci.org/publications/documents/

Liu, LM 1989, 'Identification of seasonal ARIMA models using a filtering method', *Communications in Statistics-Theory and Methods*, vol. 18, pp. 2279–2288.

Liu, Z & Todini, E 2002, 'Towards a comprehensive physically-based rainfall-runoff model', *Hydrology and Earth System Sciences*, vol. 6, pp. 859–881.

Luo, J, Wang, E, Shen, S, Zhen, H & Zhang, Y 2011, under review, Effects of conditional parameterization on performance of rainfall-runoff model regarding hydrologic non-stationarity, Submitted to *Journal of Hydrology*.

Makridakis, S & Hibo, M 2000, 'The M3-Competition: results, conclusions and implications', *International Journal of Forecasting*, vol. 16, pp. 451–476.

Moore, RJ & Bell, VA 2002, 'Incorporation of groundwater losses and well level data in rainfall-runoff models illustrated using the PDM', *Hydrology and Earth System Sciences*, vol. 6, pp. 25–38.

Moore, RJ & Clarke, RT 1981, 'A distribution function approach to rainfall-runoff modeling', *Water Resources Research*, vol. 17, pp. 1367–1382.

Moore, RJ, Cole, SJ, Bell, VA & Jones, DA 2006, 'Issues in flood forecasting: Ungauged basins, extreme floods and uncertainty', *Frontiers in flood research*, IAHS Publication vol. 305, p. 103.

Moore, RJ, Jones, DA & Black, KB 1989, 'Risk assessment and drought management in the Thames basin', *Hydrological Sciences Journal*, vol. 34, pp. 705–717.

Mulvany, TJ 1850, On the use of self-registering rain and flood gauges, *Proceedings of the Institute of Civil Engineers of Ireland*, vol. 4, no. 2, pp. 1–8.

Murray–Darling Basin Authority 2009, *River Murray drought update*, Issue 17, viewed 2 September, http://www.mdba.gov.au/media_centre/archived-information

Nash, J & Sutcliffe, J 1970, 'River flow forecasting through conceptual models part I – A discussion of principles', *Journal of Hydrology*, vol. 10, pp. 282–290.

Nash, JE 1960, 'A unit hydrograph study with particular reference to British catchments', *Proceedings of the Institution of Civil Engineers*, vol. 17, pp. 249–282.

O'Connell, PE, Nash, JE & Farrell, JP 1970, 'River flow forecasting through conceptual models part II – The Brosna catchment at Ferbane', *Journal of Hydrology*, vol. 10, pp. 317–329.

Ord, K & Lowe, S 1996, 'Automatic forecasting', *The American Statistician*, vol. 50, pp. 88–94.

Peaty, T & Seasonal Streamflow Forecasting Team 2009, *Review of the CSIRO Bayesian joint probability model for use in a seasonal streamflow forecasting service*, Water Forecasting Branch, Bureau of Meteorology, Australia.

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

Podger, G 2004, *RRL rainfall-runoff library, user guide*, Cooperative Research Centre for Catchment Hydrology, Australia.

Porter, JW 1972, 'The synthesis of continuous streamflow from climatic data by modelling with a digital computer', PhD thesis, Department of Civil Engineering, Monash University, Australia.

Porter, JW & McMahon, TA 1975, 'Application of a catchment model in southeastern Australia', *Journal of Hydrology*, vol. 24, pp. 121–134.

R Development Core Team 2011, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna viewed 2 September, http://www.R-project.org

Raupach, MR, Briggs, PR, Haverd, V, King, EA, Paget, M & Trudinger, CM 2009, *Australian Water Availability Project (AWAP): CSIRO marine and atmospheric research component: Final report for phase 3, CAWCR Technical Report No. 013*, Bureau of Meteorology & CSIRO, Australia.

Raupach, MR, Briggs, PR, Haverd, V, King, EA, Paget, M & Trudinger, CM 2011, *CSIRO AWAP run 26c historical monthly and annual model results for 1900–2009*, Centre for Australian Weather and Climate Research (Bureau of Meteorology & CSIRO), Australia.

Reggiani, P & Schellekens, J 2006, 'Rainfall-runoff modeling: Distributed models', *Encyclopedia of Hydrological Sciences*, John Wiley & Sons, Ltd.

Reilly, D 2000, 'The AUTOBOX system', *International Journal of Forecasting,* vol. 16, pp. 531–533.

Renard, B, Kavetski, D, Kuczera, G, Thyer, M & Franks, SW 2010, 'Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors', *Water Resources Research,* vol. 46, W05521.

Renard, B, Kavetski, D, Leblois, E, Thyer, M, Kuczera, G & Franks, S 2011, 'Towards a reliable decomposition of predictive uncertainty in hydrological modeling: Characterizing rainfall errors using conditional simulation', Water Resources Research, *under review*.

Rosenbrock, HH 1960, 'An automatic method for finding the greatest or least value of a function', *The Computer Journal,* vol. 3, pp. 175–184.

Ruiz, JE, Cordery, I & Sharma, A 2007, 'Forecasting streamflows in Australia using the tropical Indo-Pacific thermocline as predictor', *Journal of Hydrology,* vol. 341, pp. 156–164.

Schreider, SY, Whetton, PH, Jakema, AJ & Pittock, AB 1997, 'Runoff modelling for snow-affected catchments in the Australian alpine region, eastern Victoria', *Journal of Hydrology,* vol. 200, pp. 1–23.

Shao, Q & Li, M 2009, *A review of statistical downscaling techniques and numerical weather prediction, A report prepared for the Bureau of Meteorology,* Precipitation and Actual Evapotranspiration Products project, Water Information Research and Development Alliance (WIRADA), CSIRO Water for a Healthy Country Flagship program, Canberra.

Shao, Q & Li, M 2010, *Assessment and development of bias correction in GCM downscaling procedure,* Seasonal and Long-Term Forecasting and Prediction project, WIRADA, CSIRO Water for a Healthy Country Flagship program, Canberra.

Shao, Q & Li, M 2011a, An improved statistical analogue downscaling procedure for seasonal precipitation forecast, submitted to Stochastic Environmental Research and Risk Assessment.

Shao, Q & Li, M 2011b, A statistical downscaling procedure for forecasting monthly precipitation using POAMA output, under CSIRO internal review.

Shao, Q, Li, M, Chan, C & Wang, E 2009, *Assessing the performance of statistical downscaling techniques on POAMA to Murrumbidgee sub-catchments,* Seasonal and Long-Term Forecasting and Prediction project, WIRADA, CSIRO Water for a Healthy Country Flagship program, Canberra.

Shin, D, Schepen, A, Peatey, P, Zhou, S, MacDonald, A, Chia, T, Perkins, J & Plummer, N 2011, 'WAFARi: A new modelling system for seasonal streamflow forecasting service of the Bureau of Meteorology', *2011 International Congress on Modelling and Simulation*, Perth.

Thyer, M, Renard, B, Kavetski, D, Kuczera, G & Clark, M 2011, 'Improving hydrological model predictions by incorporating rating curve uncertainty', *34th IAHR World Congress, Engineers Australia,* Brisbane.

Thyer, M, Renard, B, Kavetski, D, Kuczera, G, Franks, SW & Srikanthan, S 2009, 'Critical evaluation of parameter consistency and predictive uncertainty in hydrological modeling: A case study using Bayesian total error analysis', *Water Resources Research* vol. 45, pp. 22.

Timbal, B, Li, Z & Fernandez, E 2009, *The Bureau of Meteorology statistical downscaling model graphical user interface: User manual and software documentation, CAWCR research report,* CAWCR (Centre for Australian Weather and Climate Research), Melbourne.

Timbal, B & McAvaney, BJ 2001, 'An analogue-based method to downscale surface air temperature: Application for Australia', *Climate Dynamics,* vol. 17, pp. 947–963.

Tuteja, NK & Cunnane, C 1999, 'A quasi physical snowmelt runoff modelling system for small catchments', *Hydrological Processes,* vol. 13, pp. 1961–1975.

Tuteja, NK, Vaze, J, Murphy, B & Beale, G 2004, *CLASS: Catchment scale multiple-land use atmosphere soil water and solute transport model*, NSW Department of Infrastructure, Planning and Natural Resources, Australia and Cooperative Research Centre for Catchment Hydrology.

Tuteja, NK, Vaze, J, Teng, J & Mutendeudzi, M 2007, 'Partitioning the effects of pine plantations and climate variability on runoff from a large catchment in southeastern Australia', *Water Resources Research,* vol. 43, W08415.

Vertessy, RA & Elsenbeer, H 1999, 'Distributed modeling of storm flow generation in an Amazonian rain forest catchment: Effects of model parameterization', *Water Resources Research,* vol. 35, pp. 2173–2187.

Wang, E, Zhang, Y, Luo, J, Chiew, FHS & Wang, QJ 2011, 'Monthly and seasonal streamflow forecasts using rainfall-runoff modeling and historical weather data', *Water Resources Research,* vol. 47, W05516.

Wang, E, Zheng, H & Luo, J 2010, 'Monthly and 3-monthly streamflow forecasting using a hydrologic model coupled with ensemble historical forcing at five catchments in southeast Australia', *Water for a Healthy Country National Research Flagship Report,* CSIRO.

Wang, E, Zheng, H, Shao, Q, Chiew, F, Luo, J & Wang, QJ 2011a, Ensemble streamflow forecasting using rainfall-runoff modelling with conditional parameterisation and GCM predictions in East Australia, in internal review for submission to Water Resources Research.

Wang, E, Zheng, H, Shao, Q & Wang, QJ 2011b, Skill improvement through conditional model parameterisation and bias correction in seasonal streamflow forecasting, submitted to the WIRADA symposium and now is under CSIRO review.

Wang, QJ & Robertson, DE 2011, 'Multisite probabilistic forecasting of seasonal flows for streams with zero value occurrences', *Water Resources Research,* vol. 47, W02546.

Wang, QJ, Robertson, DE & Chiew, FHS 2009, 'A Bayesian joint probability modeling approach for seasonal forecasting of streamflows at multiple sites', *Water Resources Research,* vol. 45, W05407.

Wigmosta, MS, Vail, LW & Lettenmaier, DP 1994, 'A distributed hydrology-vegetation model for complex terrain', *Water Resources Research,* vol. 30, pp. 1665–1679.

Wilks, DS 1995, *Statistical methods in the atmospheric sciences – an introduction,* Academic Press.

DOT6493/11

Water Information
DATA › INFORMATION › INSIGHT