Improved Modelling of the Catchments and Drainage Network in the Upper South East for Management Outcomes

Gibbs, M.S., Humphrey, G.B., Maier, H.R., Dandy, G.C

Goyder Institute for Water Research

Technical Report Series No. 15/34

GOYDER INSTITUTE FOR WATER RESEARCH

www.goyderinstitute.org

Goyder Institute for Water Research Technical Report Series ISSN: 1839-2725

The Goyder Institute for Water Research is a partnership between the South Australian Government through the Department of Environment, Water and Natural Resources, CSIRO, Flinders University, the University of Adelaide and the University of South Australia. The Institute will enhance the South Australian Government's capacity to develop and deliver science-based policy solutions in water management. It brings together the best scientists and researchers across Australia to provide expert and independent scientific advice to inform good government water policy and identify future threats and opportunities to water security.

Enquires should be addressed to: Goyder Institute for Water Research

Level 1, Torrens Building

220 Victoria Square, Adelaide, SA, 5000

tel: 08-8303 8952

e-mail: enquiries@goyderinstitute.org

Citation

Gibbs, M.S., Humphrey, G.B., Maier, H.R., Dandy, G.C., 2015, *Improved Modelling of the Catchments and Drainage Network in the Upper South East for Management Outcomes,* Goyder Institute for Water Research Technical Report Series No. 15/34, Adelaide, South Australia

Copyright

Disclaimer

Executive Summary

Predictions of streamflow a month or a season ahead provide valuable information for water resource managers for subsequent planning. This is particularly the case in catchments with a highly variable flow regime, as the streamflow is difficult to predict, and hence, manage. For the purposes of predicting streamflow, it is desirable to not only provide a value for the prediction, but also quantify the uncertainty associated with the prediction to provide an indication of the potential range of outcomes.

This report presents a range of approaches tested to provide such predictions for a study site in the South East of South Australia. Due to the flat topography and drainage infrastructure, decisions on where and when to divert flow around the landscape can be made, often with the aim to improve environmental outcomes. One of these decision points is along Drain M, which terminates at a wetland of high importance, Lake George, but water can also be diverted out of the drain to be used to support other watercourses and wetlands. With the many competing demands on this water resource, it is desirable to forecast future flows at key locations along Drain M in order to maximise the outcomes achieved from the water available.

Some form of hydrological model is required to provide predictions of upcoming streamflow in Drain M. A hydrological model captures the relationships between initial conditions, climate forcings (typically rainfall and potential evapotranspiration) and streamflow. In prediction mode, a hydrologic model calibrated with historical data is run forward in time, with input data representing forecast climate forcings, to predict the streamflow in an upcoming period. Three major factors control forecasting accuracy: (1) the ability of the hydrologic model to predict streamflow with actual forcings; (2) the accuracy of the initial conditions adopted (e.g., soil moisture and groundwater stores); and (3) the accuracy of the forecasts of the climate inputs. While the third factor is obviously important to forecast accuracy, the objective of this work is to investigate approaches to improve model predictions due to the first two factors, and adopting recently developed approaches to provide the climate forecast inputs.

A number of approaches have been considered to provide these predictions: lumped conceptual rainfall runoff models, statistical approaches, and a combination of the two. In this work, all three approaches have been considered to assess their suitability in the ephemeral catchments of the South East. Each type of model has been investigated to improve the ability to provide accurate predictions of drain flow.

For the lumped conceptual models, the use of groundwater data as another input source (along with rainfall and potential evapotranspiration) was considered. Changes in groundwater data were found to be directly related to changes in dynamic model parameters, indicating that groundwater data were indeed useful to represent changes in the catchment. However, a threshold in the relationship between model parameters and changes in groundwater level also observed, and similar results could be achieved by only considering the recent behaviour of the catchment to provide accurate simulations of upcoming streamflow. A number of investigations were also undertaken to ensure the uncertainty in the model outputs was represented reliably.

As opposed to the lumped conceptual models, the statistical models characterise system response primarily through the extraction of information contained in a set of observed data, rather than directly attempting to represent the physical processes occurring within the hydrological system. Thus, the statistical models require predictors that describe the initial catchment condition and the effect of climate during the forecast period. In addition to previous streamflows, rainfall and evapotranspiration, which are traditionally used to represent antecedent catchment conditions, a number of soil moisture datasets were considered to describe the initial catchment conditions:

- A remotely sensed soil moisture data set, produced by the European Space Agency combining both passive and actively sensed soil moisture measurements
- Modelled soil moisture from the Bureau of Meteorology's coupled ocean/atmosphere model (POAMA2) used to provide the climate forecasts
- The soil storage level modelled by the lumped conceptual models

It was found that each individual soil moisture dataset was able to provide improvement over models that did not include any soil moisture data, particularly in terms of mean forecast accuracy. Overall, it was found that the

inclusion of all three datasets was required to give the best improvement in forecast accuracy and reliability over the model which only included more traditional inputs for representing antecedent catchment conditions and did not include soil moisture data directly. As such, when developing statistical streamflow forecasting models, it is considered worthwhile to include soil moisture data obtained from large, readily available global datasets (such as climate models and remote sensing) as potential model inputs.

After the range of different models to predict streamflow in Drain M were developed, the performance of the different models was compared, both in terms of the accuracy of the average of the range of model outputs, as well as the ability to capture the uncertainty in the predicted flow. The comparison was made between: 1) the lumped conceptual models, 2) the statistical models, and 3) a combined model, where the output from the lumped conceptual model was available as an input to the statistical models. It was found that that the combined model produced the robust results across the three catchments.

As a trial of the models developed to predict streamflow, the approach was tested during the 2014 season. The lumped conceptual models were used for the trial, as the remotely sensed soil moisture data selected as an input for the combined model are not currently available in real time. However, these data are likely to become more readily available in the near future. The value of information on antecedent conditions was clear from this trial, where the output from the models dramatically improved the reliability of the forecasts compared to the rainfall forecast alone, particularly for the extremely dry months experienced in the spring of 2014.

Finally, an operational model was developed to allow for different diversion management scenarios to be assessed. This model provides the functionality to take the forecast streamflow volumes in the drainage network as an input, apply different management scenarios for the operation of Bool Lagoon, the Callendale regulator and diversions along the drain constructed as part of the REstoring FLOWS to the wetlands in the Upper South East (REFLOWS) project, and Lake George, and finally assess the likely water level in Bool Lagoon, water level and salinity in Lake George, and the volume diverted along the REFLOWS drain.

The models developed through this project are intended to provide another source of information to assist the decision making process surrounding diversions from Drain M. With forecasted volumes for the upcoming month, it is possible that the decision to divert flow can be made earlier in the season, while still ensuring that the downstream requirements of Lake George can be maintained. In turn, this results in improved use of the freshwater resource available in the region, balancing the competing environmental outcomes across the landscape.

Acknowledgements

This project was funded by the Goyder Institute for Water Research, Project E.2.4.

The authors would like to acknowledge the following data providers for access to the data necessary to enable this project to be undertaken:

- Bureau of Meteorology for access to forecasts from the Predictive Ocean Atmosphere Model for Australia (POAMA)
- Department of Environment, Water and Natural Resources for access to streamflow and groundwater data used in South Australia
- Department of Environment and Primary Industries for access to groundwater data in Victoria
- European Space Agency and Technische Universität Wien for Essential Climate Variable Soil Moisture data

The authors would also like to thank the reviewers of this report, in particular David McInerney for his constructive input to the uncertainty estimation techniques. Reviews were also provided by Kumar Savadamuthu, Department of Environment, Water and Natural Resources, and Mark DeJong, South East Water Conservation and Drainage Board.

Contents

Table of Figures

Table of Tables

1 Introduction

Predictions of streamflow a month or a season ahead are essential information required by water resource managers for subsequent planning (Wang et al., 2011). This is particularly the case in catchments with a highly variable flow regime, as the streamflow is difficult to predict, and hence, manage. For the purposes of predicting streamflow, it is desirable to not only provide a value for the prediction, but also quantify the uncertainty associated with the prediction to provide an indication of the potential range of outcomes.

This report presents the approaches developed to provide such predictions for a study site in the South East of South Australia (outlined in the following section). Two commonly used approaches were considered, conceptual and statistical rainfall – runoff models, as well as a combined approach. The models allow a seasonal rainfall forecast to be converted to a streamflow forecast, after accounting for the antecedent conditions in the catchments at the time of the forecast. The subsequent forecasts can be used to inform management decisions in the region, and identify if suitable (or excess) volumes are likely to be expected in a coming period, and in turn take full advantage of the water available.

1.1 Management Question – Drain M

The water resource management question considered in this study is a component of the drainage network in the South East of South Australia. A large proportion of the region's runoff is generated in the higher rainfall areas of the Lower South East. Historically, this runoff flowed in a northerly direction, along the watercourses adjacent to the dune ranges, parallel to the coastline. Over the past 100 years or more, these flow paths has been diverted through a series of cross-country drains, constructed to provide flood relief by draining water in a more direct south-westerly direction to the ocean. This extensive drainage network is shown in Figure 1. Drain M, which conveys water from Bool Lagoon to the ocean near Beachport (Figure 1), is the largest of the cross-country drains. This drain collects flow from a number of drainage systems located on the southern side of the drain, as denoted by the grey shaded region in Figure 1.

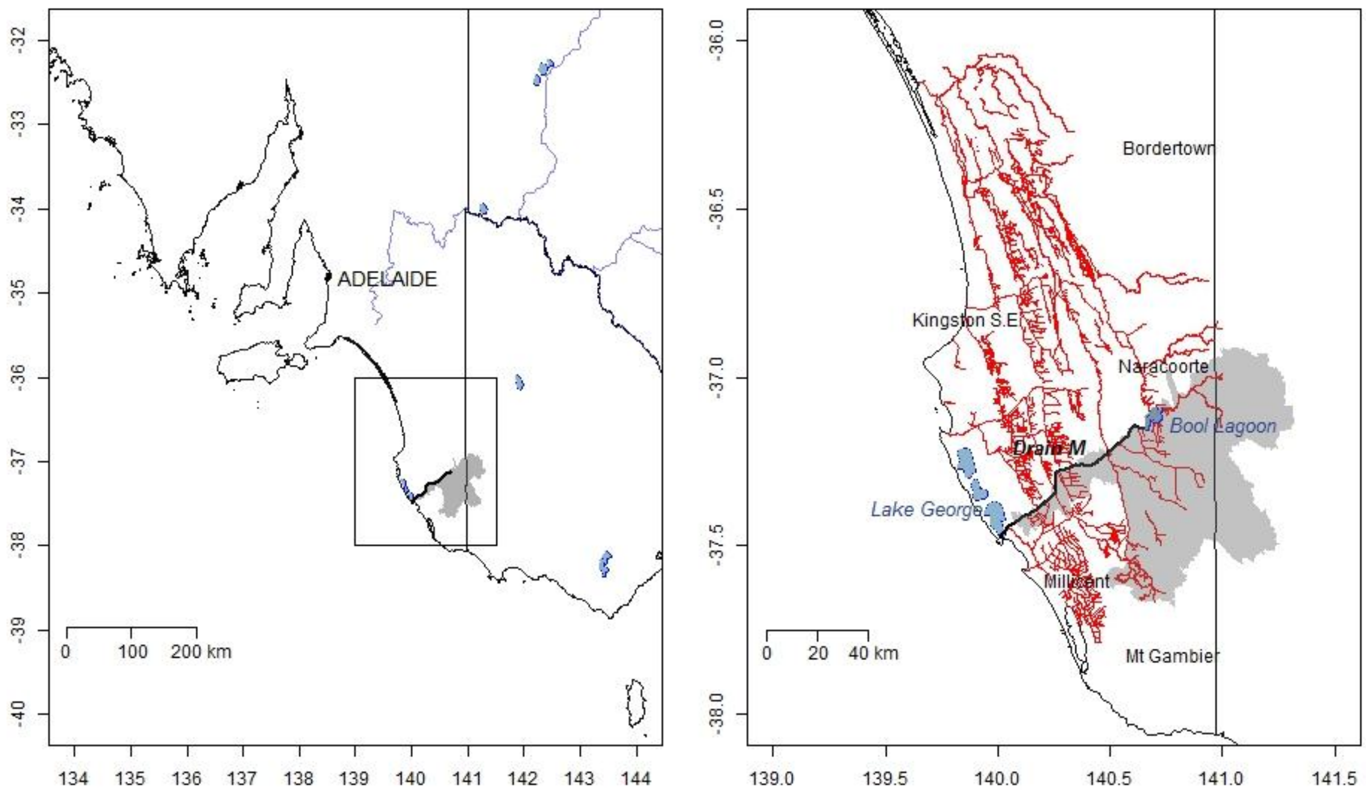


Figure 1 Map of Drainage network in the South East of South Australia

Recently, the ability to divert flow in a northward direction from this drain and restore the natural flow path to the Upper South East (USE) and the Coorong has been established. However, the termination of the Drain M system also contains a wetland of high importance, Lake George. Decisions must be made throughout the year (mainly in late winter and spring) regarding diversions from Drain M; if water should continue to drain to Lake George and the ocean, or if it should be diverted to the north, and potentially be used at a number of en route wetlands. Such decisions aim to meet the following competing objectives:

- Ensure the water requirements of Lake George are met;
- Minimise possible impacts on sea grasses, caused by fresh water flows, often with high nutrients, out to sea;
- Maximise water diverted to the north, to benefit the wetlands in the USE and potentially the Coorong.

Runoff volumes from Drain M are highly variable (between close to zero and more than is required to support Lake George, and varying over three to four orders of magnitude within a given month, as seen in Figure 2) and it is difficult to maximise the use of water available from year to year within this variability. Thus, with the many competing demands on this water resource, it is desirable to forecast future flows at key locations along Drain M in order to maximise the outcomes achieved from the water available. The modelled catchments, as well as flow, rainfall and evapotranspiration data available are outlined in Section 2 of this report.

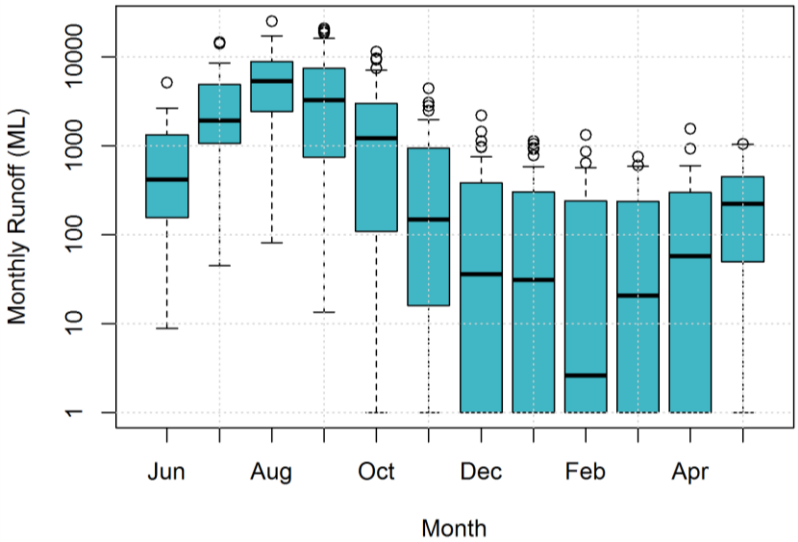


Figure 2 Variability in monthly runoff at Catchment A2390512 from 1973 - 2012

Further complicating the management of the network, the historical data may not represent future flows that could be expected under similar climatic conditions, due in part to land use change in the catchment. The change in land use has coincided with a decline in some groundwater levels in the shallow unconfined aquifer. Trends in these types of aquifers are likely to affect the streamflow response, particularly in low relief catchments where baseflow is a significant proportion of the total streamflow (such as is the case in the South East).

Petheram et al. (2011) proposed that in the low-relief, moderate rainfall catchments of southern South East Australia, relatively high groundwater levels may have amplified overland flow by reducing the storage capacity of the unsaturated zone and by facilitating organised patterns of drainage and the connection of source areas of runoff as the soil wetted up during a rainfall event. Under a falling water table, the storage capacity of the unsaturated zone is likely to be increased, and hence saturation conditions are less likely to occur, and the connectivity of source areas is likely to be less organised (Petheram et al., 2011).

Typically rainfall-runoff models are used to convert an input rainfall and potential evapotranspiration into the runoff that can be expected. However, in this case there are likely to be other changes in the catchment that will influence the amount of runoff that can be expected, such as the changes in land use and groundwater level outlined above. The potential to include extra information, in the form of groundwater levels, into a typical rainfall-runoff model is considered in Section 3.

1.2 Streamflow Forecasting Models

Some form of hydrological model is required to provide predictions of upcoming streamflow in Drain M. A hydrological model captures the relationships between initial conditions, climate forcings (typically rainfall and potential evapotranspiration) and streamflow. In prediction mode, a hydrologic model calibrated with historical data is run forward in time, with input data representing forecast climate forcings, to predict the streamflow in an upcoming period. Three major factors control forecasting accuracy (Luo et al., 2012): (1) the ability of the hydrologic model to predict streamflow with actual forcings; (2) the accuracy of the initial conditions adopted (e.g., soil moisture and groundwater stores); and (3) the accuracy of the forecasts of the climate inputs. While the third factor is obviously important to forecast accuracy, the objective of this work is to investigate approaches to improve model predictions due to the first two factors, and adopting recently developed available approaches to provide the climate forecast inputs. As such, the results from this study represent projections, i.e. given a climate forecast, what is the projected streamflow likely to be.

A number of approaches have been developed to provide these predictions: hydrological modelling approaches (Demargne et al., 2013; Wood and Schaake, 2008), statistical approaches (Bennett et al., 2014; Robertson and Wang, 2013), or a combination of the two (Robertson et al., 2013). In this work, all three approaches have been considered to assess their suitability in the ephemeral catchments of the South East. The development of the hydrological models (lumped conceptual rainfall runoff models), including uncertainty estimation, is outlined in Sections 3 and 4. However, a limitation of this approach is the need to select a specific form of the hydrological model to begin with. Statistical models remove this limitation to some extent, and the development of the statistical models is outlined in Section 5. This includes consideration of multiple data sources to represent the initial conditions in the catchment, and approaches to select only the important inputs for the development of a parsimonious model.

A comparison of accuracy and precision of the lumped conceptual and statistical models developed is undertaken in Section 6. This includes a statistical model that uses outputs from the lumped conceptual model as an input to provide the streamflow predictions, representing the combined model. The application of the models developed to predicting streamflow in Drain M in the 2014 flow season is presented in Section 5.

1.3 Water Balance Modelling

A water balance model is required to assess the impacts of different potential diversion scenarios on the wetlands that will be affected. These wetlands include Bool Lagoon, Lake George, and storage within Drain M itself. The water balance model allows water level information to be estimated, and in turn allows environmental water requirements and rules around releases from Bool Lagoon to be assessed. The development of a water balance model for the Drain M system in the eWater Source platform is outlined in Section 8.

1.4 Study Objectives

Based on the issues and questions outlined in this section, the objectives of this project are to:

- Assess the potential to include groundwater data as an input lumped conceptual rainfall-runoff models to improve the ability to simulate changes in the catchments over time (Section 3)
- Develop both lumped conceptual (Section 4) and statistical (Sections 5) models to provide predictions of streamflow in Drain M given a seasonal climate forecast
- Assess the accuracy and precision of the different model types, including a combined approach (Section 6)
- Undertake a trial of the streamflow prediction models in the 2014 season (Section 5)
- Develop a water balance model in eWater Source to allow the effect of predicted volumes and different management options on the volumes diverted and water levels in wetlands in the system (Section 8)

The following section provides an outline of the catchments in the Drain M system, and the data adopted that are common to all models developed.

2 Catchments and Data

The Drain M system is shown in Figure 3. The predominant land use is dry land pasture, and there is no major urbanisation in the catchment. The runoff from the catchment is seasonal and relatively low, with an average annual runoff of approximately 30 mm/yr. The topology of the region is flat, with mainstream slopes in the order of 0.005 m/m. The sub catchments and corresponding areas which currently contribute to flow in Drain M were defined using a 10 m LiDAR derived DEM using the ArcHydro tool, as outlined in Wood and Way (2011). The overall catchment area which currently contributes to the flow in Drain M has been divided into 3 smaller catchments, as shown by the grey shaded regions, with outlets at points A, B, and C. This division is based on the location of flow gauges as well as major management decision making locations.

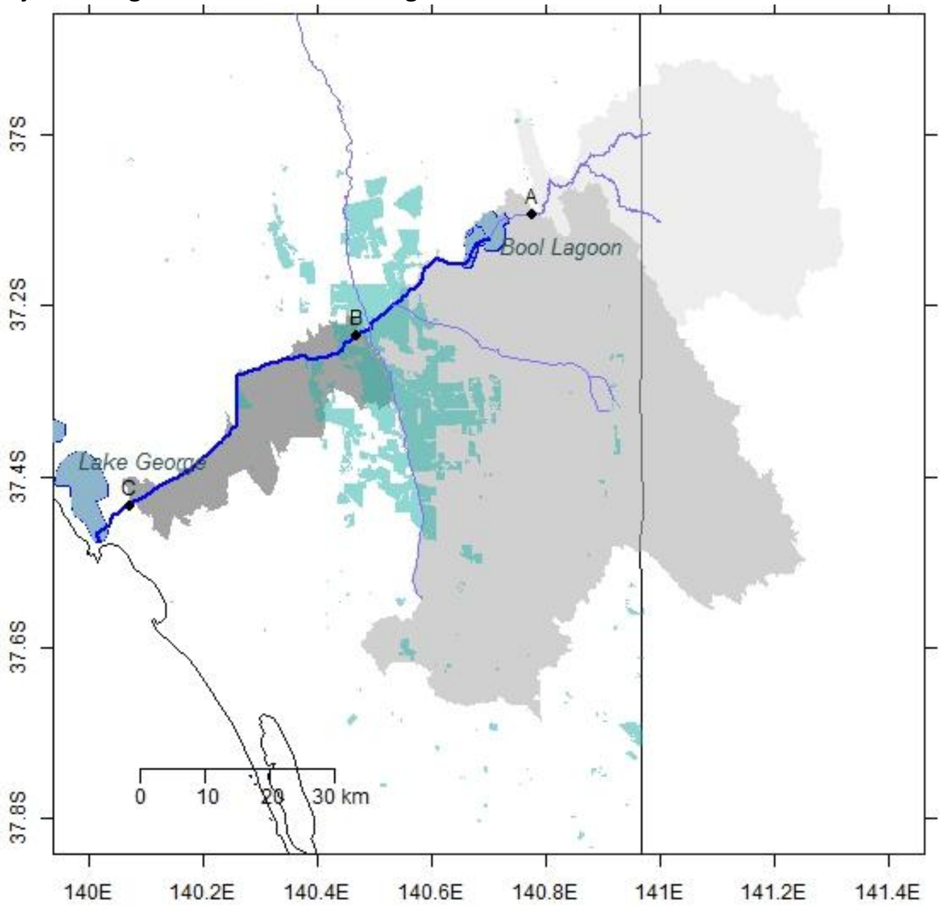


Figure 3 Modelled Drain M system, with the three sub-catchments considered outlined in shades of grey. The green shaded area represents plantation forestry in South Australia. Flow along Drain M occurs in the direction from Bool Lagoon to Lake George, before discharging to the Southern Ocean.

The most upstream catchment (to the northeast) is the Mosquito Creek catchment, which extends from Victoria into South Australia (with the border delineated by the vertical black line in Figure 3). This catchment is important for maintaining flows into the internationally significant Bool and Hacks Lagoons. The catchment consists of two major streams: Mosquito Creek which drains the southern part of the catchment; and Yelloch Creek which drains the northern part. Point A denotes the flow gauge at Struan (Site A2390519), downstream of the confluence of these streams, which records inflows to the Bool Lagoon wetlands.

Point B marks the location of the flow gauge just downstream of the Callendale regulator (Site A2390514), which is used to control flows in Drain M and provides the potential to divert flows northward through the floodway along the Bakers Range watercourse (constructed through the Restoring flows to the Upper South East (REFLOWS) project) and into the USE. The catchment between points A and B includes flows released from Bool and Hacks Lagoons, which, as well as being significant wetlands, act as major water storages in this region. As such, only minor releases from the lagoons toward point B have occurred in the past 15 years. The flow to Point B also includes flow from the Southern Bakers Range drain, which runs along the catchment's western boundary. The southern part of this catchment can experience high rainfall and, in the past, the Bakers Range drain has contributed major flows into Drain M. However, Blue Gum forestry plantations (denoted by the green shaded regions in Figure 3), which have been established in the region since 1998, are expected to have altered the hydrology in this sub-catchment over the past decade, reducing flows in the Bakers Range drain. Furthermore, Drain C, which connects into Drain M part way between Bool Lagoon and Callendale and primarily drains the south-eastern cross-border portion of the catchment, rarely produces runoff as a result of the very flat landscape, depression storage, and the introduction of centre point irrigation and plantation forestry.

The most downstream catchment between points B and C denotes the local catchment downstream of Callendale to where Drain M terminates at Lake George. All three catchments shown in Figure 3 contribute to flow in Drain M at point C, which marks the location of the flow gauge at Woakwine (Site A2390512), near to the drain's terminal point.

The three catchments will henceforth be referred to as A2390519, A2390514 and A2390512, after the flow gauges which record their respective outflows. Details of the catchment areas, annual rainfalls and annual flows are given in Table 1. Datasets common to all components of this report, flow, rainfall and evapotranspiration data, are presented in the remainder of this section. Other datasets considered in this work are presented in the section where they are first used.

Table 1 Catchment area and mean annual rainfall and flow. Mean rainfall calculated over the period 1970 – 2012, and flow dependent on the data available. Mean Annual flow was calculated from the total flow recorded at that location, however Area represents the sub catchment area only (e.g. dark grey area in Figure 3 for catchment A2390512)

Catchment	Area (km^2)	Mean Annual Rainfall (mm)	Mean Annual Flow (GL)
A2390519 (A)	1003	606	22.29
A2390514 (B)	2200	667	29.27
A2390512 (C)	383	676	50.06

2.1 Flow

Daily flow data are available at locations A (site A2390519, Mosquito Creek at Struan), B (site A2390514, Drain M at D/S Callendale Regulator) and C (site A2390512, Drain M at Woakwine Amtd 5.1km) from 1971 until 2012. Additionally, flows have been recorded at the outlet of Bool Lagoon (site A2390541, Drain M at D/S Bool Lagoon Outlet) from 1985 until present (28 years). The locations of these flow gauges are shown in Figure 4.

Structures used to relate recorded water level to streamflow for the three catchments considered are fixed concrete weirs constructed for the purposes of streamflow measurement. Both A2390512 and A2390519 stations have v-notch cross-sections to improve sensitivity at low flows, where A2390514 has flat cross sections and therefore could be expected to be less sensitive at low flows. All stations are free flowing downstream and are constructed drains upstream with a constant slope and cross section, which would be expected to provide a reliable relationship between stage and discharge. Gauging records for the four stations are extensive, with between 78 and 166 flow gaugings at each station, with the 90^{th} percentile deviation 10.5% for stations A2390512 and A2390519, and 16.3% at station A2390514. Given the regular cross sections and high number of gaugings available to develop stage-discharge relationships, output data uncertainty is expected to be low for the stations considered.

As can be seen in Figure 5, flows in Drain M predominantly occur between June and November, while for the remainder of the year there is typically no (or only very low) flow. Time series plots of the monthly flows at each of the gauges are shown in Figure 6. As can be seen, there is a significant decrease in monthly flow volumes for each catchment after the mid to late 1990s. This is particularly noticeable for gauge A2390541 (which is affected by storage in Bool Lagoon), where it can be seen that there have only been two releases from Bool Lagoon after this time. This is in part due to the millennium drought, where rainfall was substantially below average for 2006-2009. A plot of cumulative flow against cumulative rainfall from the period the record started until 2012 is presented in Figure 7, which would be expected to be relatively linear in the case of no changes in the catchment and high quality data. It can be seen that there was change in the relationship between rainfall and runoff, particularly in the

A2390519 and A2390514 catchments, occurring around the late 1990s. This rainfall and flow data are expected to be of high quality, and as such this change in relationship implies a change in the catchment rainfall-runoff relationship occurred around this time.

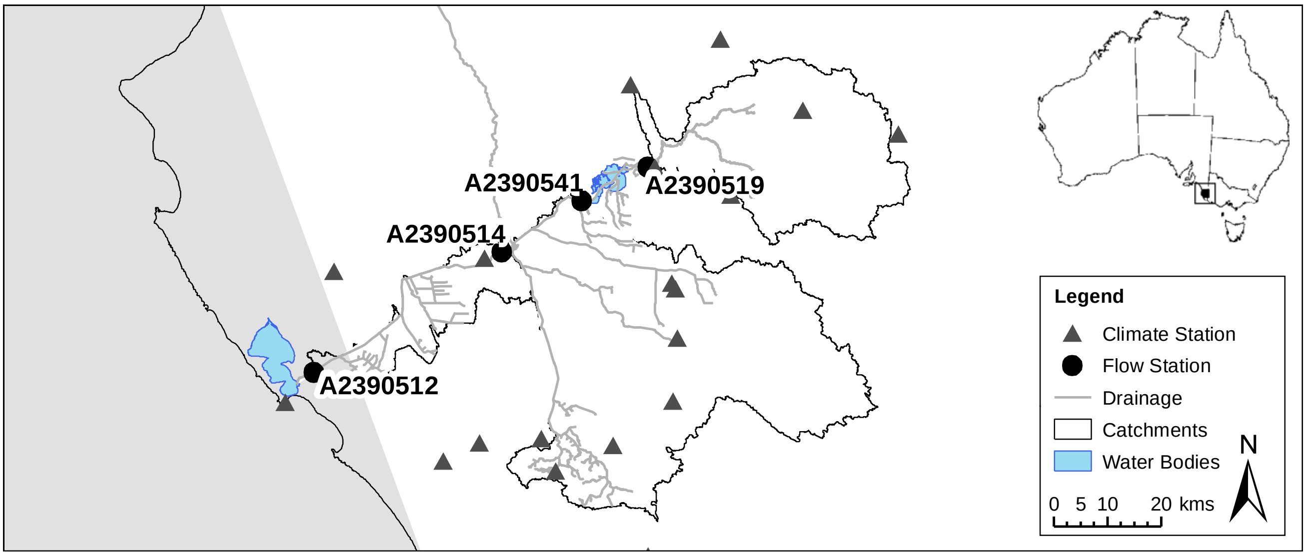


Figure 4 Catchments considered and flow stations. The diversion point is at station A2390514, where flow can be directed to the North

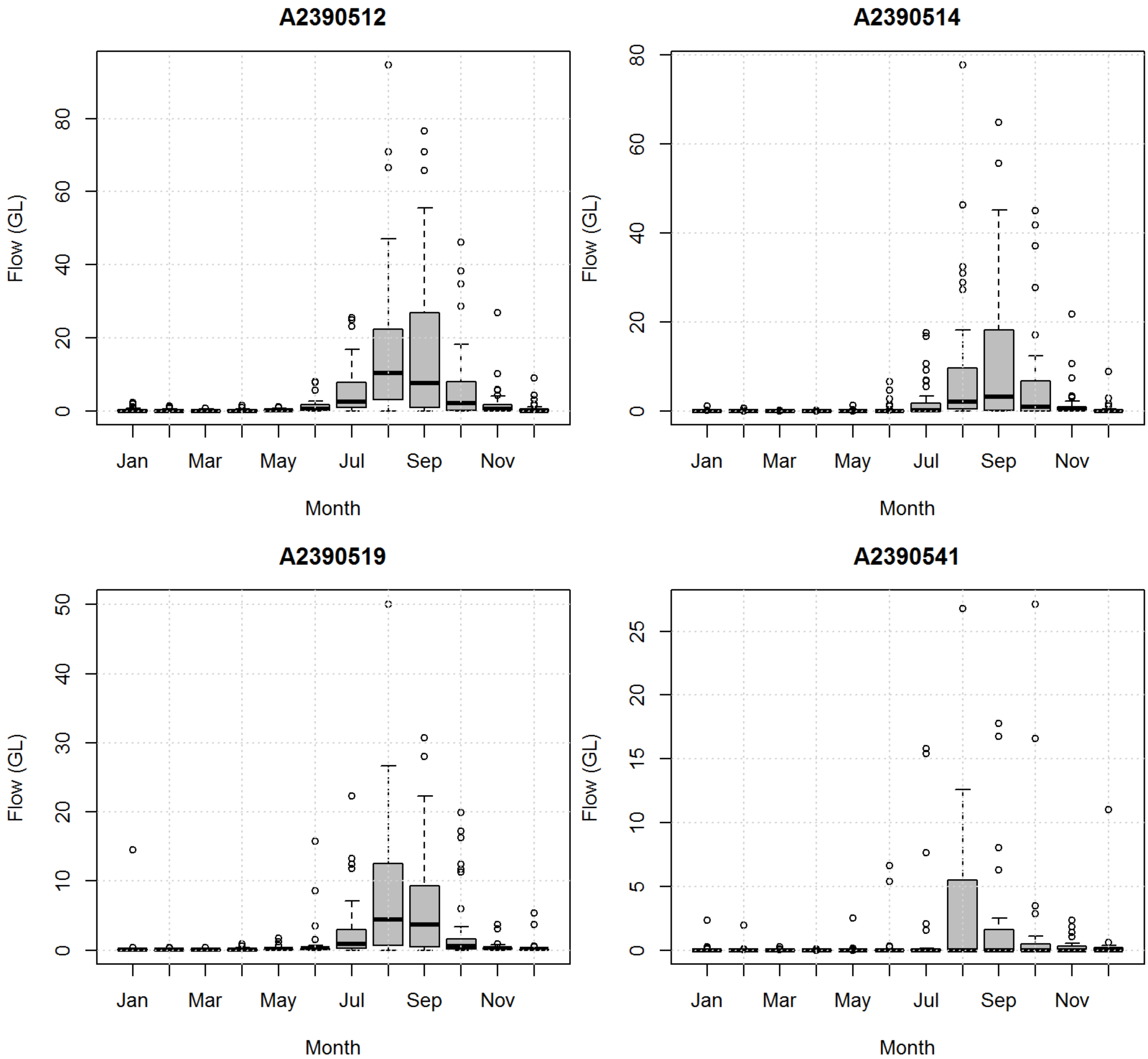


Figure 5 Boxplots of total monthly flow for each gauge along the Drain M system

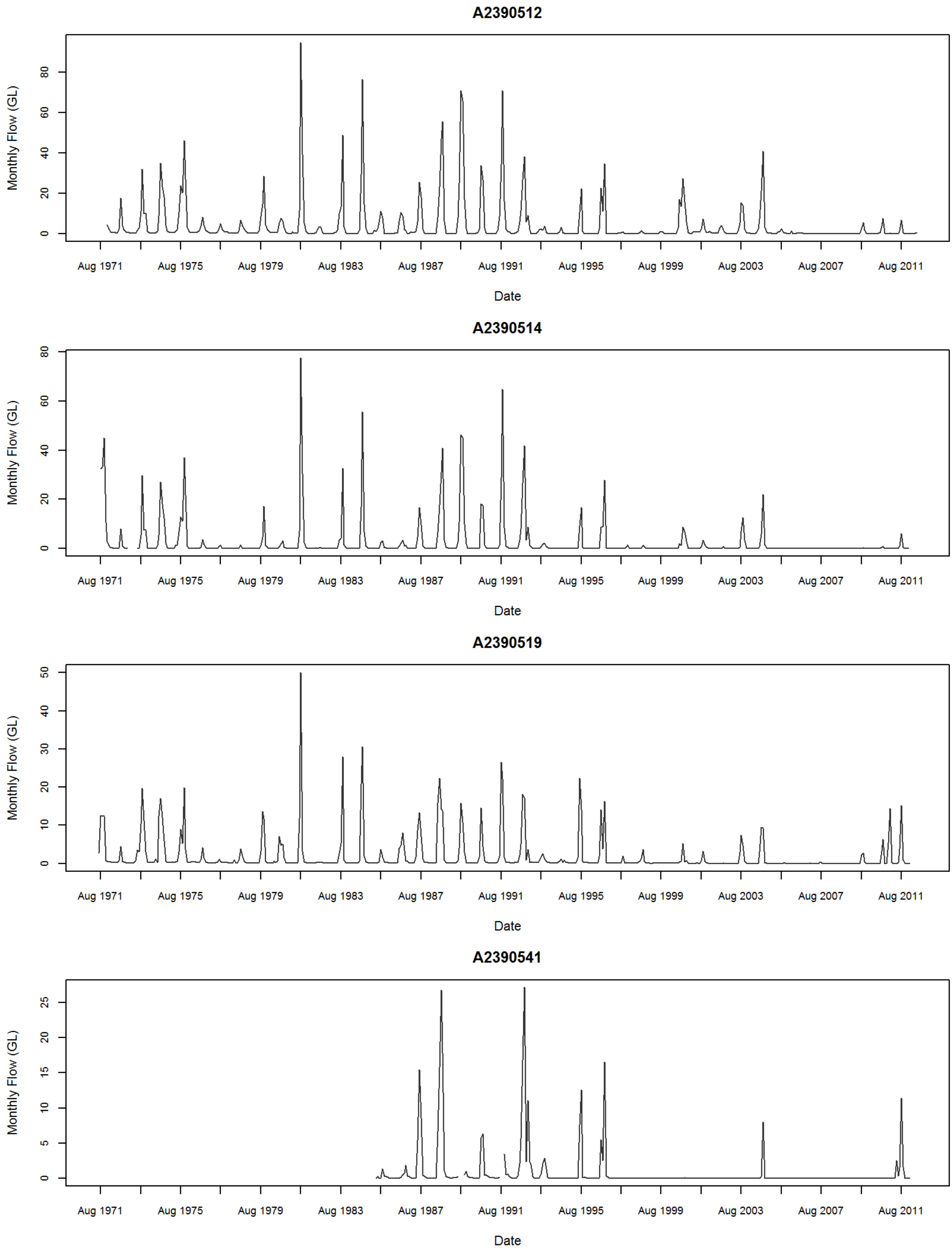


Figure 6 Monthly flow time series at each of the flow gauges

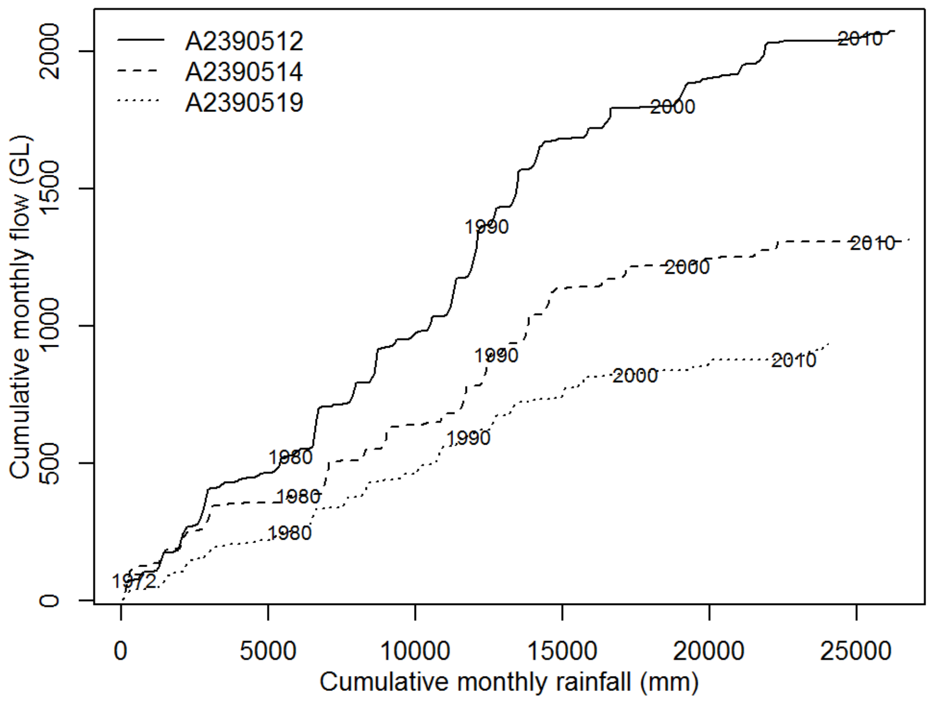


Figure 7 Cumulative Rainfall vs cumulative runoff for the three catchments

2.2 Rainfall and Evapotranspiration

The mean annual rainfall for the region is approximately 600 - 675 mm and the mean annual FAO56 potential evapotranspiration (PET) (Allen et al., 1998) is approximately 1000 mm. The SILO Patched Point Dataset (Jeffrey et al., 2001) was used for the rainfall and FAO56 evapotranspiration data adopted, with the stations used shown in Figure 4. A Thiessen polygon approach was used to combine stations and produce one time series each of rainfall and evapotranspiration for each catchment. Boxplots of the resulting total monthly rainfall and evapotranspiration can be seen in Figures Figure 8 Figure 9, respectively, for the three catchments. As can be seen, the highest rainfalls are experienced in the winter months, with rainfall exceeding evapotranspiration in the months May – September.

To test for possible errors or changes, the input data were compared to high quality data available from the Bureau of Meteorology for each catchment and each variable. For the catchment averaged rainfall input:

- Rainfall for catchment A2390512 was compared to the rainfall data at Robe (station 26026, 40 km north),
- Rainfall for catchment A2390514 was compared to the rainfall data at Penola (station 26025, within the catchment), and
- Rainfall for catchment A2390519 was compared to the rainfall data at Natimuck (station 79036, on the eastern edge of the catchment)

Each of the stations used for comparison are part of Australia's high-quality gauge network (Lavery et al., 1992). To test for input data errors in the PET, temperature data (one of the main inputs to Penman–Monteith FAO56 equation) were compared to the catchment average temperature for each station. For the temperature data:

- Temperature for catchment A2390512 was compared to the temperature data at Robe (station 26026, 40 km north),
- Temperature for catchment A2390519 was compared to the temperature data at Nhill (station 78015, 60 km north),
- No high quality data were available close to the A2390514 catchment.

Each temperature dataset used for comparison is part of the Bureau of Meteorology Australian Climate Observations Reference Network – Surface Air Temperature network (Trewin, 2013). To undertake the comparisons, the Homogeneity testing approach of Wijngaard et al. (2003) was adopted. No statistically significant evidence of inhomogeneity was found, implying that the record retained is sufficient for the intended analysis, and reduces the potential contribution of poor data quality to predictive uncertainty.

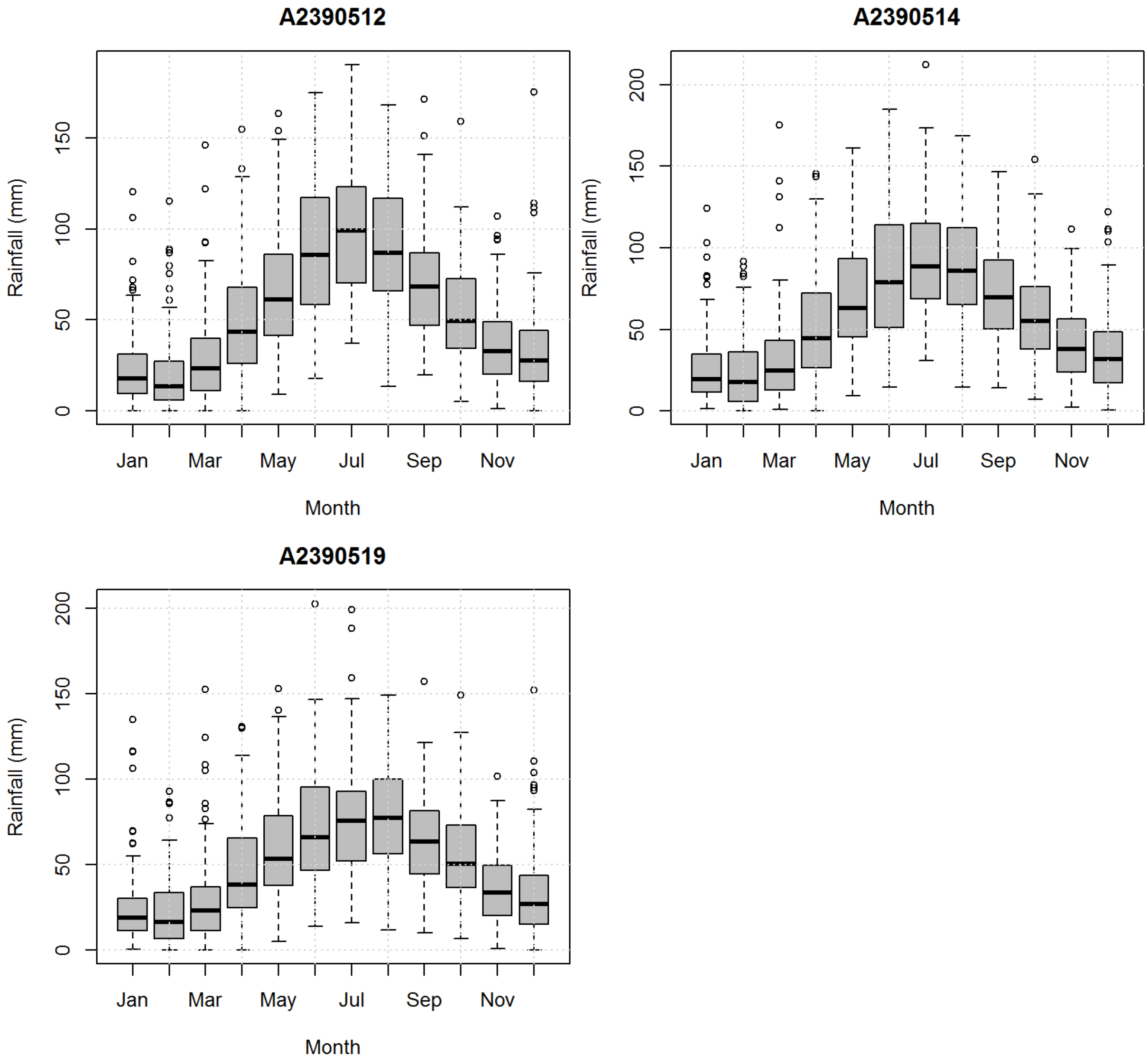


Figure 8 Boxplots of monthly rainfall, averaged across each catchment

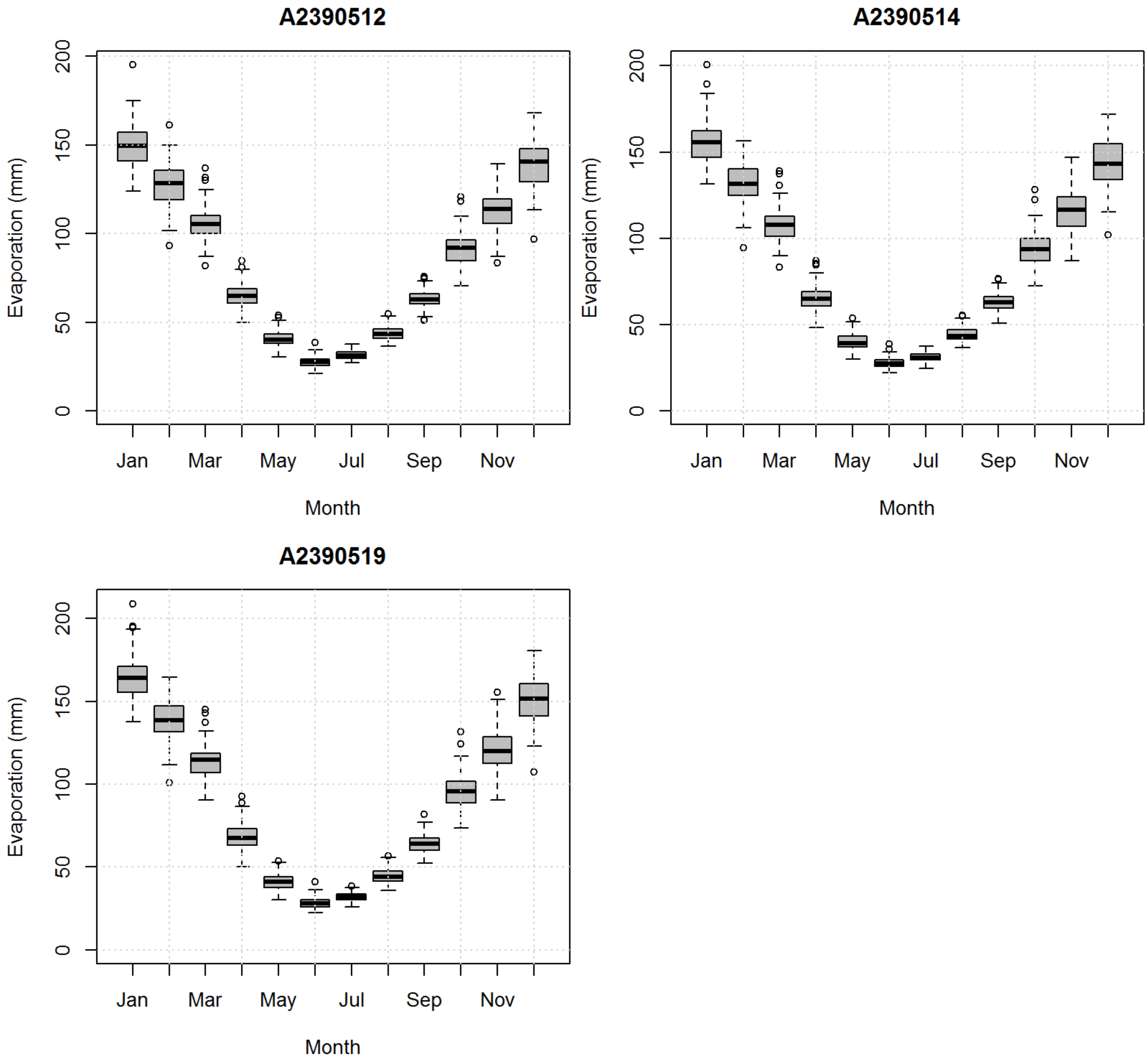


Figure 9 Boxplots of monthly evapotranspiration, averaged across each catchment

3 Potential to use Groundwater Data in Lumped conceptual Catchment Models

There is a shallow unconfined aquifer underlying most of the South East of South Australia, and the connection to this aquifer can play a large role in the response of runoff to rainfall. But, the most common inputs to lumped conceptual rainfall-runoff (CRR) models are only time series of rainfall and potential evapotranspiration (PET). Based on these inputs, it is unlikely that CRR models have any information to be able to represent changes in streamflow due to changes in groundwater level. If the changes in groundwater level are expected to be due to extended periods of below average rainfall, it is plausible that this may be captured in the rainfall time series. However, it is difficult for the storages in a simple CRR models to represent long term processes such as this, and if the cause of the decline in groundwater levels is not rainfall alone, there is no model input to drive a change in the simulated hydrologic regime.

Groundwater data are another source of data that may be able to be used as an input to provide useful information on changes in a catchment that influence the rainfall runoff relationship, that is not captured by the typical inputs of rainfall and PET. Hence, the objective of this section is to assess the potential to include groundwater data as an input lumped conceptual rainfall-runoff models to improve the ability to simulate changes in the catchments over time. In this section two studies were undertaken to investigate the potential to improve catchment modelling through the use of groundwater data as a model input:

1. The first study recalibrated the model parameters to shorter-overlapping periods to investigate changes in model parameters over time, and the co-variation of model parameters with a trend in groundwater data, without prescribing certain relationships.
2. Based on the outcomes from the first study, the second study explicitly develops a relationship between model parameters and groundwater data. In this study, the model structure was also modified in an attempt to improve the ability to represent such changes.

3.1 Impact of Variable Parameters on Model Performance

In this section, a study is presented that was designed to test if calibrated parameter values for the standard GR4J model vary over time. If this is the case, there is potential for extra input data to be used to help represent this variability, which is considered in Section 3.2.

3.1.1 Methodology

Catchment A2390519 has been considered in this initial study due to the long high quality data record, and clearly observed change in the rainfall-runoff relationship (Figure 7). The GR4J model (Perrin et al., 2003) was selected as the base CRR, as it is a parsimonious model that explicitly accounts for non-conservative (or 'leaky') catchments, and has demonstrated good performance for Australian conditions (Coron et al., 2012). Karstification of limestone is perhaps the most widespread reason worldwide for non-conservative catchments (Le Moine et al., 2007), and this characteristic describes the shallow unconfined aquifer in the catchment considered. The model schematic can be seen in Figure 10, and further details of the model structure and parameters (X1, X2, X3 and X4) can be found in Perrin et al. (2003).

Model calibration was undertaken using the hydromad R package (version 0.9-15) (Andrews et al., 2011). The SCE algorithm was used to calibrate the models with parameter relationships adopted from Duan *et al.* (1994b) with 10 complexes and a maximum of 10,000 evaluations for the same objective function as previously adopted. Data from days with poor quality code were ignored in the calibration, and a warm-up period of one year was used. Each year started on the first of March to represent the water year.

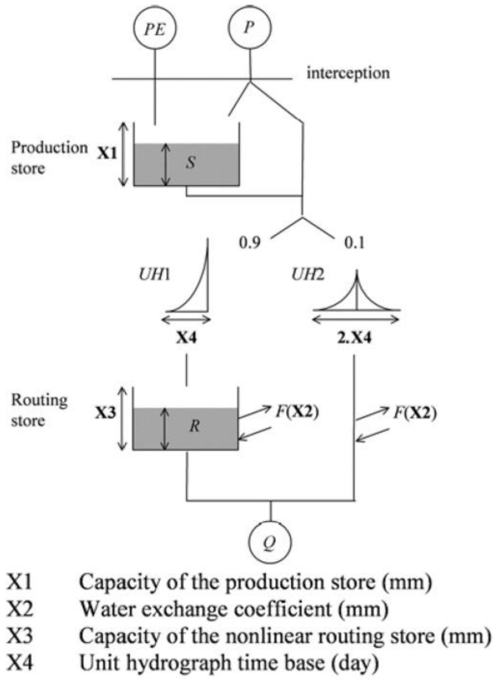


Figure 10 GR4J Structure (Perrin et al., 2003)

A number of calibration scenarios have been undertaken to investigate how the GR4J model parameters change over time to represent the change in hydrologic regime for the catchment considered:

1. The model was calibrated to the period with groundwater data available, starting in 1987. This first step was to assess how well the model can represent the observed data in a "typical" application,
2. A rolling seven year period was used to calibrate the model parameters to identify those parameters that change over time to represent the change in hydrologic regime,
3. The parameters observed to change over time were calibrated individually over the seven year rolling period, to reduce the effect of any correlation with other parameters. For this case, the values for the three parameters not involved in the calibration scenario were taken as those identified in calibration scenario 1.

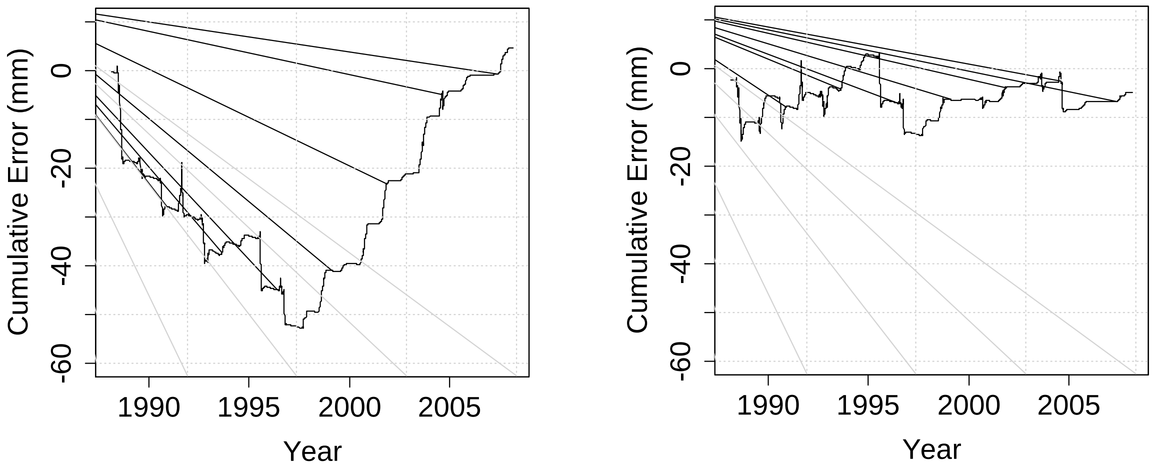
The time varying parameter values from Scenario 3 were compared to the catchment average change in groundwater level, to assess the relationship between individual parameter values using groundwater levels without prescribing a certain functional form.

3.1.2 Results

The cumulative error in the daily flows after calibrating all four GR4J model parameters (calibration scenario 1) is presented in Figure 11(a), where negative values indicate that the model underestimated the observed flow. The effect of the change in the hydrologic regime on the model errors can be seen in Figure 11(a), where to represent the overall volume accurately, the model generally under predicted the observed flow before 1996, and then generally over predicted the flow after this time.

By recalibrating the model parameters each year the X3 parameter varied over time, allowing the model to represent the change in the water balance over time more accurately than for the case where all parameter values were stationary. The cumulative error from the models where X3 only has been calibrated can be seen in Figure 11(b), where the cumulative error in the modelled results can be seen to be much closer to zero for the whole period compared to Figure 11(a). Figure 11(b) has been produced by combining the simulated flow values for the rolling period calibration results, where the simulated flows for the middle year of the seven year period have been extracted and combined to produce a time series. This constructed time series is likely to have discontinuities in the storage levels in the model when switching from one model to the next, and may result in errors in the overall water balance. However, the result does provide an indication of how the model may behave with a variable value for X3, and it can be seen that the models are able to represent the runoff volume relatively accurately over the whole period considered, as opposed to the constant model parameter results presented in Figure 11(a).

For calibration scenario 2, the value calibrated for the time base of unit hydrograph in scenario 1, X4 = 4.7 days, has been fixed for this analysis, as this parameter would only be expected to influence the timing of the simulated peaks as opposed to the volume simulated, and would not be expected to change unless there were changes to the efficiency of the creeks in conveying the flow downstream (e.g. clearing of vegetation). The results from calibrating the remaining three parameters over a seven year rolling period for calibration scenario 2 are presented in Figure 12. The value for each year in Figure 12 has been derived from the middle of the rolling period, for example the value for 2000 represents the model calibrated to the period 1/3/1997 to 28/2/2004, after a one year warm-up period starting on 1/3/1996. To allow the trend in the parameter values to be compared in Figure 12, the ranges adopted for the y axis are ±50% of the mean value for each parameter. Two values for X2 are greater than this range, -11.1 mm in 2000 and -9.5 mm in 2001. While there may be a small trend in X1, this parameter was found to be largely stationary over time. Both X2 and X3 were found to have declining trends, and as such are investigated further. For the GR4J model, a negative value for the X2 parameter indicates the export of water from the catchment, conceptually as a groundwater exchange.

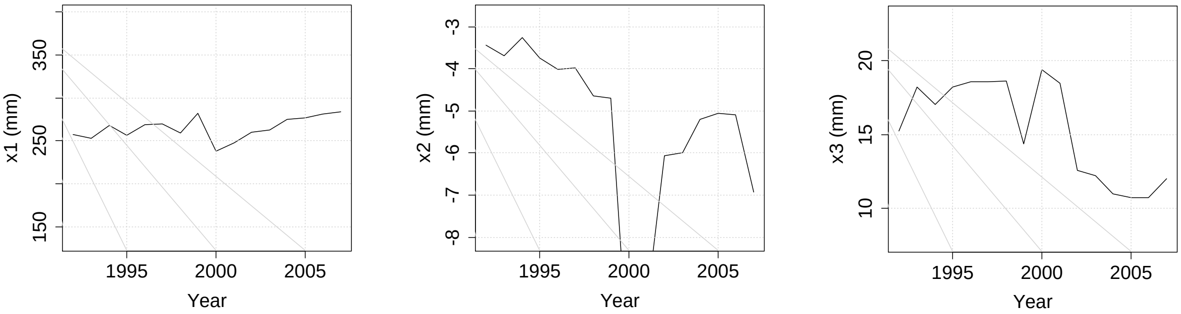


(a) (b)

Figure 11 Cumulative error for with constant values for all four GR4J parameters (a) and allowing X3 to vary over time (b)

For calibration scenario 3, both X2 and X3 have been calibrated individually to remove the effect of any interaction between the two parameters on the values calibrated. For this case, the values identified from calibration scenario 1 have been adopted for the parameters that are not involved in the calibration process, X1 = 276.1 mm, and depending on the variable being investigated, X2= -3.9 mm or X3 = 14.9 mm. The resulting values from the X2 and X3 analysis were highly correlated (with R=0.96), as can be seen in Figure 13.

The resulting trend in X3, while keeping the three other parameters constant, is presented in Figure 14. The results for parameter X3 are presented, as the trend in X2 is more erratic (as seen in Figure 12(b) for the calibration trial including all three parameters), and given the strong correlation between X2 and X3 in Figure 13, a similar result would be expected for both parameters. The difference between Figure 14 and Figure 12(c) is due the calibration process used, as all parameters were calibrated concurrently in Figure 12, but only X3 was calibrated in Figure 14, with the other parameter values held constant.



(a) (b) (c)

Figure 12 Change in three of the four GR4J parameters over time

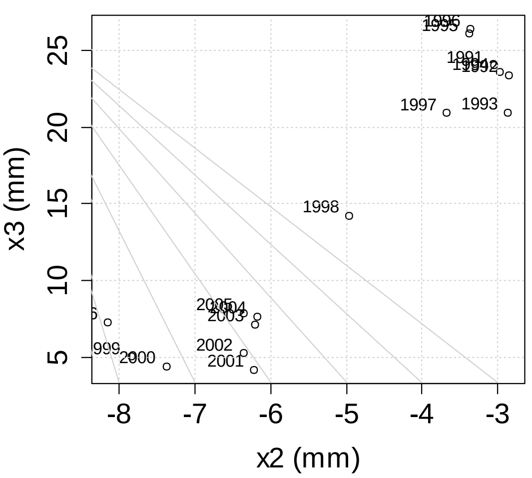


Figure 13 Correlation between X2 and X3

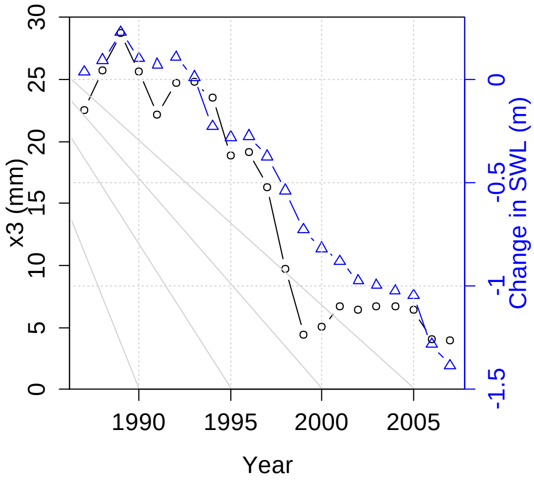


Figure 14 Relationship between X3 and change in groundwater standing water level (SWL)

The blue line in Figure 14 is the catchment average change in groundwater level, derived from the six wells identified in the catchment in the riparian zone (within 1 km of the streams). Very good agreement between the calibrated X3 values and the change in groundwater level relative to the average level prior to 1995 can be seen in Figure 14, suggesting that there is the potential to allow X3 (or X2, given the strong correlation between the two parameters) to vary over time, driven by a relationship with groundwater level.

3.1.3 Discussion

In the GR4J model, the groundwater exchange term, F, is given by Equation. 1:

$$F = X2\left(\frac{R}{X3}\right)^{7/2} \tag{1}$$

This term acts on flow from both the production and routing stores. In this work, the calibrated values for X2 were negative, representing water export from the system. As such, it is reasonable for X2 and X3 to change as the runoff coefficient decreases, as both a larger value (or more negative in this case) for X2 or smaller value for X3 will result in larger values for F. The high correlation between X2 and X3 presented in Figure 13 is also expected given the calculation of groundwater exchange in Equation 1, however X3 is also involved in the calculation of flow out of the routing store.

Figure 13 suggests that a state change in the catchment behaviour may have occurred, with high X3 (or low X2) up until 1998, and lower values after this time. This change corresponds to a 0.5 m drop in the catchment average groundwater level (Figure 14). As such, variable parameter values may not be necessary, and this new state in the catchment may be represented by the updated parameter values. It is unclear if a future period of extended wet conditions, resulting in more recharge and higher groundwater levels, would result in the hydrologic regime returning to that observed previously, or if this is unlikely to occur due to land use/practice changes and abstraction of the groundwater resource. The answer to this question may determine if time varying parameter values are necessary or not.

3.1.4 Summary

For the catchment considered, stationary parameter values were found to underestimate the flow before the change in runoff response had been identified, and overestimated the flow after this time. By adopting a rolling seven year period to calibrate the parameters for the GR4J model considered, the two parameters involved in the groundwater exchange term were both found to vary over time. Very good agreement between the calibrated parameter values involved in the groundwater exchange term (particularly X3) and the change in catchment averaged groundwater level, relative to the average level prior to 1995, was found, indicating that there is the potential to inform time varying lumped conceptual rainfall runoff model parameters using groundwater data. This is investigated further in the following section.

3.2 Incorporating groundwater data into lumped conceptual rainfall runoff models

In this study, a number of modifications to a standard CRR model have been considered to improve the ability to represent long term trends in the rainfall-runoff relationship. For this section, the Reedy Creek - Mt. Hope Drain catchment (A2390513) was considered as the third catchment, rather than A2390514, due to the limited availability of groundwater data in the A2390514 catchment, and high quality streamflow and groundwater data in the A2390513 catchment. The location of this catchment can be seen as the most southerly catchment in Figure 15. The models considered, the data used for calibration and validation of the models and the approaches used to calibrate the parameters are outlined in the following section.

3.2.1 Methodology

3.2.1.1 Groundwater Data

To investigate the changes in groundwater level in the catchment, groundwater wells within 1 km of the streams were identified, with an example of Mosquito Creek presented in Figure 16. The 1 km range was selected to represent a riparian zone influencing streamflow, with the 1 km distance identified to produce consistent trends in the groundwater levels. The six wells identified are located in a shallow unconfined aquifer, and data from the wells have been combined to produce the catchment average change in groundwater level. The observations (generally taken quarterly) have been averaged to produce an annual time series, as long term trends are of most interest in this work, as opposed to the seasonal cycle or recharge and decline in the shallow aquifer. The average groundwater level from when the record started at each well up to 1995 has been calculated and subtracted from the annual time series to standardise the water level recorded at each well. Finally, data from the six wells each year were averaged to produce a catchment averaged groundwater trend dataset on an annual time scale. Data for these wells were extracted from SA (https://www.waterconnect.sa.gov.au/GD/Pages/default.aspx) and Victoria databases (http://www.vicwaterdata.net/vicwaterdata/home.aspx), as the border between the states is located between wells JOA013 and B105672 in Figure 16. The resulting trend for the A2390512 catchment can be seen in Figure 17.

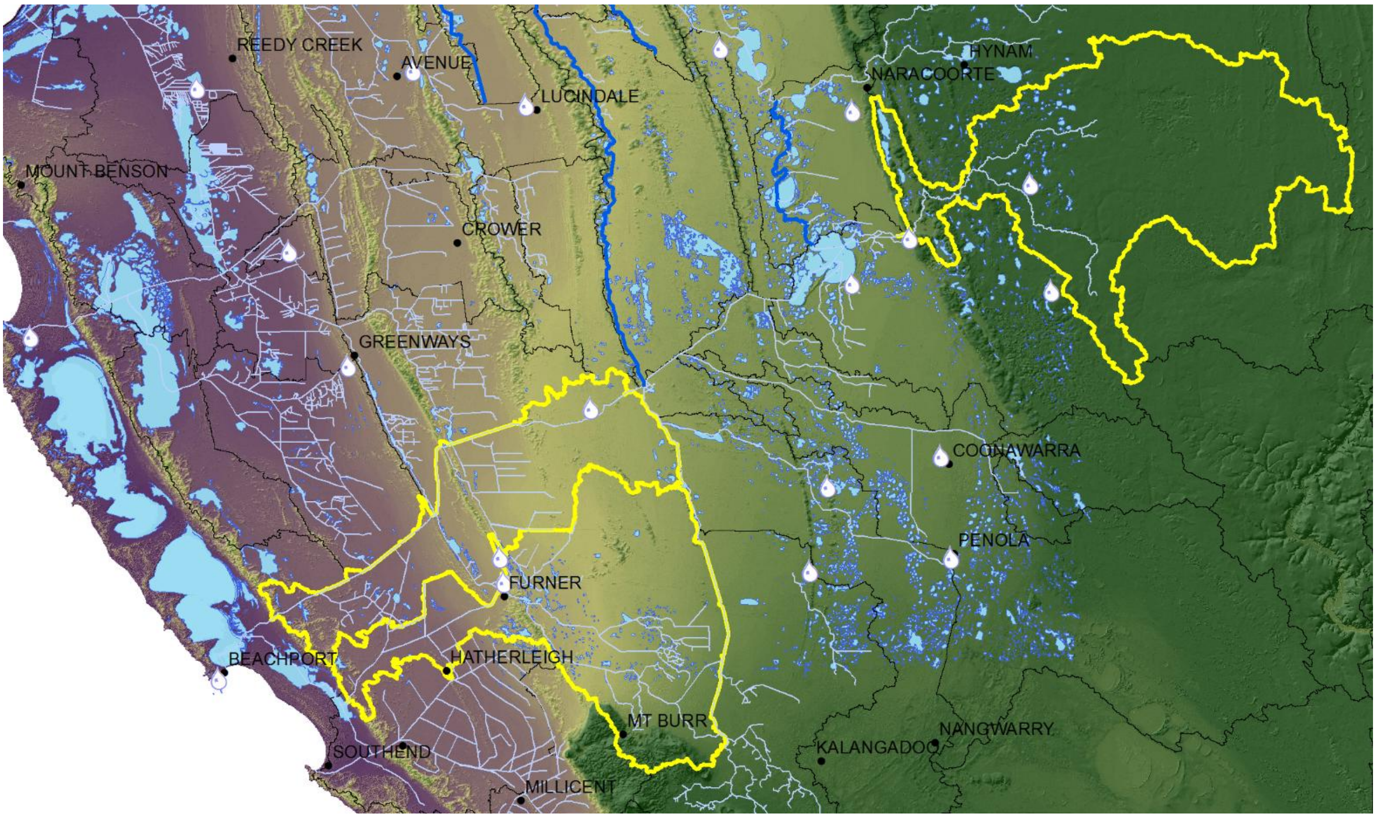


Figure 15 Catchments considered for the groundwater data investigation.

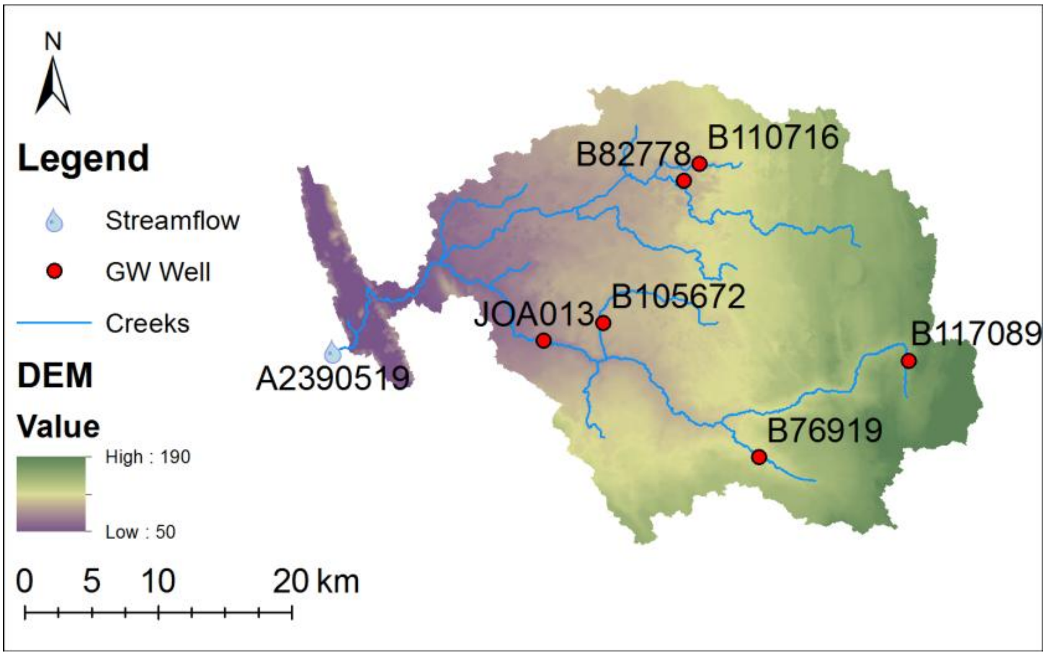


Figure 16 Location of groundwater wells used in the Mosquito Creek catchment

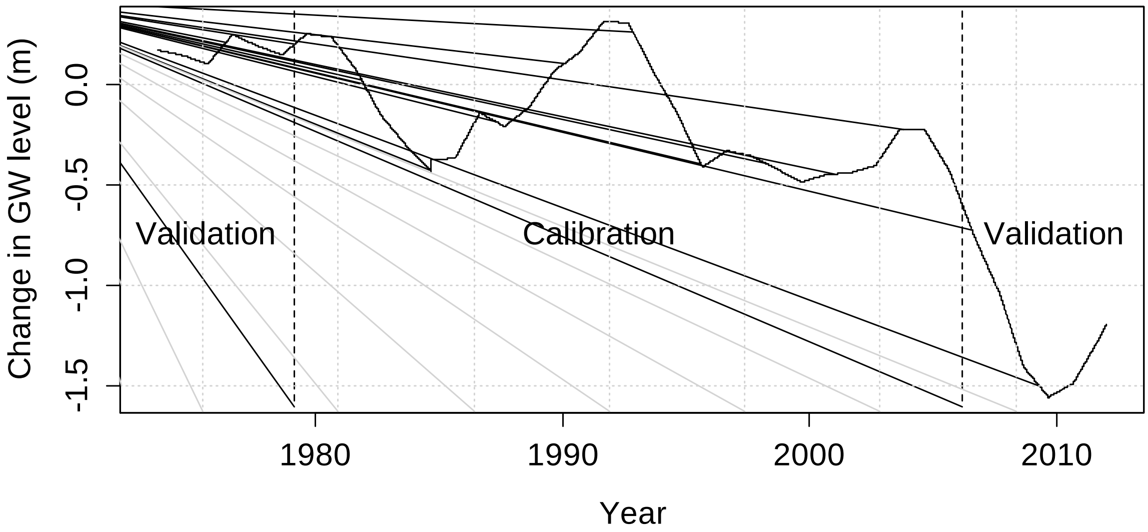


Figure 17 Trend in GW levels based on wells in the riparian zone for catchment A2390512. Vertical dashed lines delineate the calibration and validation periods.

Given that the change in the catchment average groundwater level coincides with the change in the rainfall runoff relationship in the mid 1990s (Figure 7), it is proposed to investigate if the optimal parameters of a lumped conceptual rainfall runoff model that represent the rainfall runoff relationship also change around this time. If this is the case, there is the potential for the groundwater level data to be used to inform rainfall runoff model parameters and represent this change in hydrologic regime dynamically.

3.2.1.2 Models

A modified version of GR4J has also been considered, where there is an initial store as part of the nonlinear routing store that must be filled before there is any discharge to the stream. The size of this initial store is a calibration parameter, T, where T<X3. An increase in this threshold will result in less runoff.

Seven different model configurations have been considered, as outlined in Table 2. The first, M1, is the standard four parameter GR4J model. The second, M2, is a version of GR4J with seven parameters but no capability for time varying parameter values, by including scaling factors in the rainfall and PET time series to account for any systematic errors in the input data, and to calibrate the split between the direct runoff and routed runoff, specified as a value of 0.9 in the original specification of GR4J. These scaling factors have been introduced where necessary to provide a fair comparison between model configurations, so that the different configurations have the same number of free parameters to calibrate. Compared to standard GR4J (M1), M3 adds a threshold on the routing store, T, before flow is released from this store, and allows for a comparison of the benefit of this change to the model structure by comparing to M1. M4 is similar to M3, but also includes the split and scaling factors on the PET data to increase the number of parameters to seven. The PET scaling factor can be considered analogous to a calibrated catchment averaged crop factor applied to the FAO56 PET data. In models M5 and M6, parameters X1 or X2 are varied with the groundwater data, and a PET scaling factor is included to again have 7 parameters. Finally M7 is similar to M3, but treats the threshold level in the routing store, T, as dependent on the groundwater data, requiring seven parameters.

Table 2 Summary of the model structures considered

Configuration	Calibrate Threshold	Calibrate Split	Scaling Factors	Groundwater covariate	Number of Parameters
M1	No	No	-	-	4
M2	No	Yes	P, ET	-	7
M3	Yes	No	-	-	5
M4	Yes	Yes	ET	-	7
M5	No	No	ET	X1	7
M6	No	No	ET	X2	7
M7	Yes	No	-	T	7

To allow the model parameters to change with the change in groundwater level time series (for models M5, M6 and M7), three parameters have been introduced to replace one of the original GR4J parameters. The variation with the change in groundwater level data was introduced into the model by replacing the original GR4J parameter, X (i.e. X1, X2 or T), with:

$$X = \begin{cases} X_a & if\ GW > GW_a \\ X_b & if\ GW < GW_b \\ X_b + (X_b - X_a)(GW - GW_b)/GW_b & otherwise \end{cases} \tag{2}$$

where:

- GW is the change in groundwater level time series, with negative values representing a decline in the historical (pre 1995) groundwater level (e.g Figure 17),
- X_a representing one extreme of the parameter value, corresponding with a change in groundwater level of GW_a meters, and
- X_b the other extreme of the parameter value, occurring at a change in groundwater level of GW_b meters.

Through initial testing, this threshold behaviour was found to be necessary to determine a suitable relationship between the model parameters and the change in groundwater level. For changes in the groundwater level between GW_a and GW_b, the parameter value is linearly interpolated between X_a and X_b. In this work, GW_a has been set to zero, representing no change in the historical average groundwater level, and GW_b is a calibration parameter. Based on this, two more parameter values are introduced to the standard GR4J configuration by incorporating groundwater co-variation (i.e. one parameter is replaced with three parameters).

Three model configurations that use groundwater data have been considered, with the X1, X2 and T parameters changing with the groundwater data. There are two processes in GR4J that remove rainfall from the water balance before it is simulated as runoff: evapotranspiration and inter-catchment transfer. The first model considered implemented a relationship between X1 and the groundwater data. This model varies the amount of actual evapotranspiration, as the rate is determined as a function of the level in the production store and the size of the production store (Perrin et al., 2003):

$$E_S = \frac{S\left(2 - \frac{S}{X1}\right)\tanh\left(\frac{En}{X1}\right)}{1 + \left(1 - \frac{S}{X1}\right)\tanh\left(\frac{En}{X1}\right)} \tag{3}$$

where:

- E_s = actual evapotranspiration rate
- E_n = net evapotranspiration capacity
- S = level in production store

As such, changing X1 is likely to change the actual evapotranspiration calculated for a given net evapotranspiration rate.

The second model configuration varied X2 with the trend in groundwater data. X2 is the loss rate, and is used in GR4J by:

$$F = X2\left(\frac{R}{X3}\right)^{7/2} \tag{4}$$

where R is the level in the routing store. A change in X2 will directly change F, which is then subtracted from the flow generated from both the routing and production stores.

The third model configuration varied the threshold storage (T) for release from the routing store with the trend in groundwater data. In this case, a minimum storage within the routing store is required before the flow spills out as streamflow, however below this level, water in the routing store can still be lost from the catchment as inter-catchment transfer. This process can be conceptualised as a dead storage in the flow path (including soil storage), which must "wet up" before streamflow is simulated at the bottom of the catchment. For the case where T can vary with the change in groundwater level, the threshold can vary over time, representing the increased volume required to simulate streamflow when the groundwater level is lower.

3.2.1.3 Calibration and Validation Periods

The data available have been split into three periods for each catchment, two five year validation periods, one at the start and the other at the end of the record, with a longer calibration period in the middle, as shown in Figure 17. This approach was taken to test the model performance on two potentially different rainfall – runoff relationships, to determine if the model can represent any changes that have occurred.

For all lumped conceptual rainfall runoff models, the model is calibrated to the streamflow generated within the catchment. This means that the flow into the catchment was subtracted from the flow recorded at the bottom of the catchment, to only consider the flow that was generated within the subcatchment boundary. For example, flow at A2390514 was subtracted from the flow at A2390512, to calibrate to the remainder generated within the A2390512 catchment. As travel times are typically less than a day within the subcatchments, the recorded data were simply subtracted from one another.

From Table 3 it can be seen that the two validation periods are slightly drier than the calibration period, but not by a large amount (less than 6% in all cases). However, the runoff coefficients (average annual runoff / average annual rainfall) are lower in the validation periods than the calibration periods, less than half in the most recent validation period. Gaugings and discussions with hydrographers in the region suggest that the changes are not likely to be due to data errors at the streamflow gauge, and are expected to represent an actual change in the catchment runoff response.

Table 3 Catchment information and periods used for calibration and validation

Station Number	A2390512	A2390513	A2390519
Area (km^2)	382.8	538.0	1001.6
Validation Period 1	1/9/1974 – 1/3/1979	1/9/1974 – 1/3/1979	1/9/1975 – 1/3/1980
Precipitation (mm/yr)	620	707	573
Runoff Coefficient (%)	9.9	3.3	2.7
Calibration Period	1/3/1979 – 1/3/2006	1/3/1979 – 1/3/2006	1/3/1980 – 1/3/2006
Precipitation (mm/yr)	659	712	587
Runoff Coefficient (%)	12.2	6.0	4.0
Validation Period 2	1/3/2006 – 31/12/2011	1/3/2006 – 1/10/2011	1/3/2006 – 1/9/2011
Precipitation (mm/yr)	626	684	578
Runoff Coefficient (%)	1.8	2.8	1.7

There is likely to be some information in the rainfall time series to explain some of these changes, for example a change in the distribution of rainfall to more consistent compared to more extreme events generating runoff. However, it is unlikely that this is the only cause of the change in runoff coefficients, and in this case more information about the trends occurring in the catchments is likely to be useful in modelling the runoff generated.

3.2.1.4 Model Calibration

The objective function used to calibrate the model parameters for each model configuration applied to each catchment was similar to the default function used in the time series processor (TSPROC) for the Parameter Estimation (PEST) software (White et al., 2014):

$$\mathrm{OF} = \left(R^2\left(\sqrt{Q_d}, \sqrt{X_d}\right) + R^2(Q_m, X_m) + R^2(\log\left(Q_{FDC}\right))\right)/3 \quad (5)$$

Q_d is the daily flow time series, with a square root transform applied to emphasize the broader flow regime, Q_m are the monthly volumes, Q_{FDC} is the flow duration curve, calculated as the flow exceeded at five percentile intervals.

The R^2 statistic is used to balance the three objectives by normalising by the observed variance, which will all tend toward one as the metric improves. Adopting different time periods for the different components reduces the mathematical relationship between the components, which should be considered to ensure that the combined function is balanced (Węglarczyk, 1998). The inclusions of volumetric and exceedance-time characteristics in the calibration process can promote the estimation of a realistic set of parameter values (Westenbroek et al., 2012). Also, for many modelling applications, it is crucial that a model predict exceedance-time characteristics as accurately as possible under future climatic or management conditions (Westenbroek et al., 2012).

Three optimisation algorithms have been used to calibrate the models based on this objective function, the Shuffled Complex Evolution (SCE) (Duan et al., 1994a), Differential Evolution (DE) (Mullen et al., 2010) and the Dynamically Dimensioned Search algorithm (DDS) (Tolson and Shoemaker, 2007). Each algorithm was used to calibrate each of the seven model configurations on each of the three catchments 10 times, to assess how reliable each optimisation algorithm was in finding near optimal solutions. The stopping criterion adopted was 10 000 evaluations of the fitness function.

The calibration algorithms and model configurations were implemented in a modified version of the Hydrological Model Assessment and Development (hydromad) R package (package version 0.9.16 with R 2.15.1) (Andrews et al., 2011). Default values were adopted for most parameters of each optimisation algorithm. 20 complexes were specified for the SCE, with the other parameter value adopted based on the relationships provided by Duan et al. (1994b), as well as a population size of 100 for the DE. The DDS algorithm does not have any parameters that need to be specified by the user. None of these algorithm parameters were calibrated based on the performance of the algorithm on the catchment calibration problem, and as such are intended to provide a fair comparison between the algorithms. However, as the algorithms have not been tailored to the problem, it is difficult to identify the best performing algorithm from the calibration results, for example a different combination of parameters may change the ranking of the algorithms.

3.2.2 Results

Results from the model calibration study can be seen in Figure 18. In the following section, a comparison of the results from the different optimisation algorithms is undertaken, before a comparison of the different model structures, both with and without time varying model parameters.

3.2.2.1 Optimisation Results

For the comparison of the optimisation algorithms, only the results for the calibration period have been considered. Model performance for the validation periods is more relevant to the model configuration comparison, and is considered in the following section. 10,000 evaluations were found to be sufficient for the Evolutionary Algorithms to converge to a single solution, with the majority of runs converging in close to half of this number of function evaluations.

It can be seen from the black points on Figure 18 that there was very little difference between the calibration algorithms. Of the 63 comparisons that can be made (SCE compared to DDS, DDS compared to DE, and DE compared to SCE, for 3 catchments and 7 model configurations), 56 of the median values (from the 10 runs for each comparison scenario) were within 1% of each other, as well as 47 of the maximum values. These very similar results from different search algorithms provide more confidence that near optimal solutions are being identified for the sets of model parameters.

Within these small differences, the SCE tended to produce the best results, where of the 21 cases (3 catchments x 7 models), the algorithm found the best solution, and had the best median solution 15 times. For the median solution, the DE found the best solution four times, with the remaining two found by the DDS algorithm. For the maximum, the DE found the best solution for one case, in the remaining five cases the DDS algorithm found the best solution. The results from the SCE algorithm are summarised (as minimum, median and maximum values across the 10 calibration runs) in Table 4 for each catchment and calibration or validation period.

As noted above, no effort has been made to improve algorithm performance, for example DE performance may improve with a different mix of population size and number of generations to make up the 10,000 evaluations. However, it is noted that the DDS algorithm has no parameters to calibrate. The calibration results for fewer evaluations have not been investigated, where it could be the case that one algorithm converges to the solution found quicker than the others.

3.2.2.2 Model Comparison Results

First, the four models that do not include groundwater dependent parameter values based on trends in groundwater information are compared, before a comparison of the three models that do have variable parameter values.

3.2.2.2.1 Models with Stationary Parameters

For this case, models M1 – M4 are compared, representing standard GR4J, GR4J with the threshold storage in the routing store, and both models with the split between direct and routed flow calibrated, as well as scaling factors for the PET input, and in the case of M2 rainfall input.

There was little difference between the models on the calibration period across the catchments. As might be expected, slightly higher objective function values were obtained by the models with the most parameters, M4 for catchments A2309512 and A2390513, and M2 for catchment A23906519.

For the first validation period, M4 performed the best for catchments A2390512 and A3290519, where even the worst result from the 10 runs better than the best result for the other models with stationary parameters (with some exceptions for A3295012). When compared to the performance of M2, this result indicates that adding the threshold on the GR4J production store allowed for better performance than a similar model that did not have this functionality, but had the same seven parameters. However, comparing M1 and M3 for these catchments and first validation period, adding the threshold parameter only to GR4J was detrimental to performance, suggesting that this change to the model structure alone was not sufficient to improve the model performance for this case. In comparison, for the catchment that did not have a trend in the rainfall-runoff relationship, A2390513, the two models with the threshold parameter (M3 and M4) performed the worst, with M2 providing the best performance for validation period 1.

Validation period 2 represented the most recent data considered, commencing during a very dry period. For this period, the magnitude of the objective function values were much lower than for the calibration period or validation period 1 for catchments A2390512 and A12390519. For these catchments, M4, which included the threshold to commence flow and all seven parameters produced substantially better results, indicating this extra process was beneficial to model performance for a validation period that included a very dry period.

The model performance on validation period 2 was more consistent for A2390513, and again model M2 provided the best performance. This result indicates that for this catchment the threshold process was detrimental to representing the rainfall – runoff relationship for this catchment, even though higher objective function values were obtained for M4 on the calibration period for this catchment.

For catchments A2390512 and A2390519, the best performance for both validation periods from the models that do not use groundwater data was M4, generally followed by M2 (for the best model found by the DE algorithm, but the variability across calibration runs indicates that some models had lower objective function values than M1 and M3). In general, the improved performance from M2 and M4 compared to M1 and M3 for the validation period indicates that introducing the scaling factors on model inputs has not lead to the models being overfitted, and including the extra parameters still produced better model performance on data unseen in the model calibration period.

3.2.2.2.2 Models with Groundwater-dependent Parameter Values

As was the case for the models with stationary parameters, there was little difference between the models on the calibration period across the catchments. The exception to this was the M7 model for catchment A2390512, which had substantially higher calibration results than M5 and M6, and the highest median result and best solution across all models.

The calibration period performance for the three variable parameter models (M5, M6 and M7) was similar to M2 for catchment A390513, the validation performance was worse in all cases. Similarly to the previous results for including the threshold storage for catchment A2390513, the validation performance indicates that there was also no benefit in allowing a parameter value to change with the groundwater data for this catchment.

For catchments A2390519 and A2390512, comparing the three models that were informed with the groundwater data, generally M6 that included a variable X2 parameter produced the best results; in terms of both calibration and validation objective function values, as well as the variability in objective function values from one calibration run to another. For catchment A2390512, on the first validation period M7, followed by M6, produced the best results. However, on the second validation period, M6 resulted in by far the best results, significantly better than M5. In comparison, M7 had highly variable results, resulting in negative median objective function values.

For catchment A2390519, the M6 model resulted in the best model for the calibration and first validation periods, with the worst solution found over the 10 calibration runs better than the best solution found for M5 and M7 across the 10 runs. However, generally the differences can be seen to be small. For the second validation period, M7 and then M6 produced by far the best results on this driest period, again demonstrating the value of using groundwater dependent parameter values for this period in particular.

3.2.2.2.3 Overall Performance

Overall, for catchment A2390513, it can be seen that model M2 produced the best results across the three periods. This catchment had no real trend in the rainfall – runoff relationship, and as such there was no benefit in adding the threshold flow process or co-varying with model parameters with the trend in groundwater levels. As the performance for M2 is better than M1 for both validation periods, it suggests that calibrating scaling factors for the rainfall and PET inputs, as well as calibrating the split in the flows improved the model performance.

For the catchments that did have a change in the rainfall runoff relationship, A2390512 and A2390519, adding the threshold process did improve model performance for the models that did not have groundwater dependent parameter values, particularly on the second validation period. However, the improvement in the objective function found was even greater for a varying X2 parameter values based on the trend in groundwater levels, with M6 providing better objective function values than M2 for all time periods for catchment A2390519. For some periods (validation period 1 for catchment A2390512 and validation period 2 for catchment A2390519) the M7 model produced the best results, however the performance of this model was much more variable across the 10 calibration runs, indicating it is more difficult to calibration and less robust that M6, which provided similar median performance, but much more consistently.

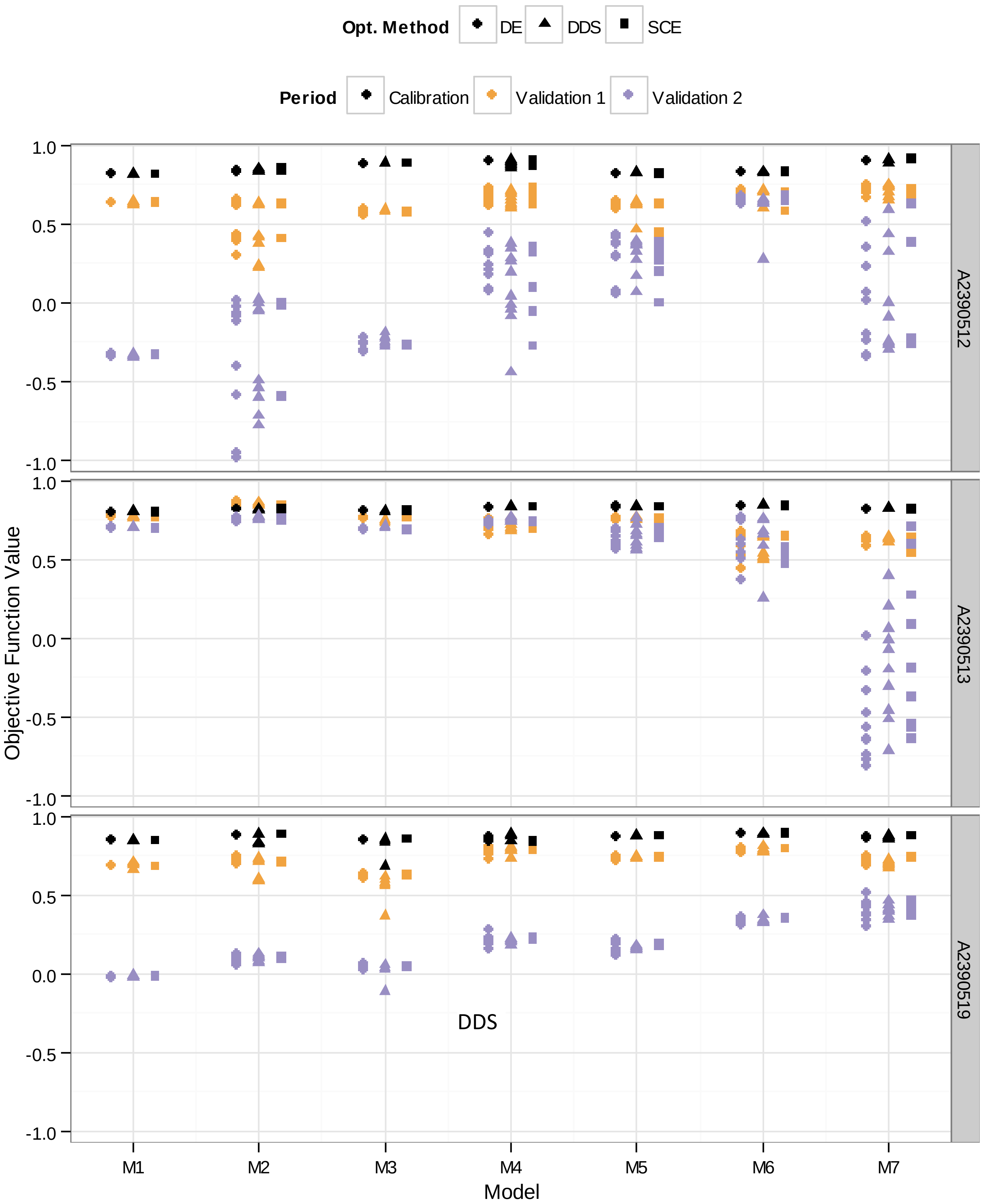


Figure 18 Calibration results, with the different optimisation algorithms shown as different shapes for each model (M1-M7), and the different time periods for calibration and validation as different colours.

Table 4 Model comparison results for the different calibration and validation periods and catchments. Values presented are for the SCE algorithm only, and the shading from low (red) to high (green) are scaled for each period separately.

Catchment	Model	Calibration Period			Validation Period 1			Validation Period 2		
		Minimum	Median	Maximum	Minimum	Median	Maximum	Minimum	Median	Maximum
A2390512	M1	0.820	0.820	0.820	0.639	0.639	0.639	-0.332	-0.331	-0.329
	M2	0.843	0.843	0.853	0.414	0.626	0.629	-0.589	-0.010	-0.002
	M3	0.891	0.891	0.891	0.582	0.583	0.583	-0.262	-0.261	-0.261
	M4	0.873	0.893	0.912	0.625	0.693	0.732	-0.271	0.323	0.359
	M5	0.826	0.827	0.827	0.446	0.628	0.629	0.004	0.355	0.393
	M6	0.831	0.835	0.835	0.587	0.704	0.707	0.654	0.660	0.683
	M7	0.911	0.911	0.915	0.693	0.728	0.728	-0.254	-0.240	0.630
A2390513	M1	0.806	0.806	0.806	0.774	0.777	0.778	0.700	0.703	0.707
	M2	0.822	0.822	0.822	0.846	0.848	0.851	0.753	0.755	0.760
	M3	0.812	0.812	0.812	0.767	0.767	0.769	0.693	0.694	0.695
	M4	0.840	0.840	0.840	0.700	0.707	0.714	0.741	0.742	0.744
	M5	0.840	0.840	0.840	0.766	0.769	0.770	0.634	0.660	0.707
	M6	0.845	0.845	0.845	0.646	0.647	0.649	0.477	0.540	0.581
	M7	0.825	0.826	0.829	0.549	0.639	0.639	-0.631	-0.047	0.717
A2390519	M1	0.850	0.850	0.850	0.687	0.687	0.687	-0.011	-0.010	-0.010
	M2	0.889	0.889	0.890	0.706	0.711	0.712	0.098	0.105	0.111
	M3	0.859	0.859	0.859	0.625	0.628	0.629	0.047	0.048	0.049
	M4	0.843	0.843	0.843	0.787	0.790	0.792	0.217	0.228	0.232
	M5	0.880	0.880	0.880	0.736	0.741	0.744	0.177	0.186	0.192
	M6	0.895	0.895	0.895	0.798	0.798	0.799	0.352	0.358	0.359
	M7	0.879	0.880	0.880	0.739	0.745	0.746	0.377	0.394	0.473

3.2.3 Discussion

Much of the variability in the second validation period for M7 can be attributed to extrapolation of the relationship controlling the groundwater dependent parameter values, where the groundwater level changes much more in the validation period than it did in the calibration period (e.g. Figure 17). However, this was also observed for other models structures that did not include variable parameters, for example M4 for A2390512, where the calibration results are very similar, but there is large variation in the validation results. This is likely due to equifinality issues arising from introducing more parameters, where multiple combinations of model parameters can produce similar results. For example, for the DE results, the correlation between the X2 and PET multiplier is R=0.89 (Figure 19). Both parameters lose water from the water balance, through either increased loss or increased actual ET, and as such there was a high correlation between the two parameter values across the 10 calibration runs.

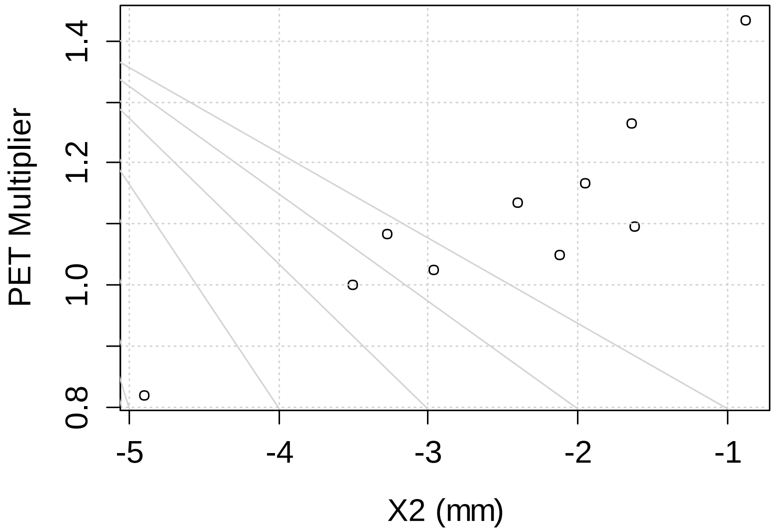


Figure 19 Correlation between X2 and PET scaling factor for DE solutions for M4 on catchment A2390512. The objective function value found for these parameter sets was between 0.902 and 0.907 for all 10 runs.

M2 generally outperformed M1 (this was not the case for the first validation period for some individual models), on both calibration and validation data. As such, the inclusion of scaling factors on rainfall and/or PET and calibrating the split between the direct runoff and routed runoff improved model performance, even in validation. Further work is required to separate the effects of the three parameters, to determine if one is more influential than the others.

Care is required when producing a suitable time series for the groundwater level at the catchment scale, and this can be relatively labour and data intensive compared to the other data inputs used for rainfall-runoff models. Data combined from numerous wells should represent the same physical aquifer, as there may be different trends in different aquifers that are not connected. The 1 km riparian zone adopted in this work to select wells has appeared to produce consistent trends in the groundwater levels; however the sensitivity of the results to this distance has not been tested and is likely to be catchment, particularly soil type, specific. It may also be desirable to represent the within year variation in groundwater levels, particularly for aquifers such as this where there is a seasonal trend of recharge and decline each year. There is a need for some standardisation to combine the data from a number of wells, as the depth to groundwater across the catchment was between seven and 13 m below the surface, and averaging these values is likely to mask trends in the wells closer to the surface, which are likely to be of most interest to the connection with the streams.

3.2.4 Summary

The results presented in this section have demonstrated that varying CRR parameters with groundwater data is beneficial in cases where a change in the rainfall-runoff relationship has been identified, and as such validation performance is very different to calibration performance. For example, the objective function for the validation period was very similar to the calibration period for catchment A2390513, and in this case there was no benefit in including the extra model parameters. However, for the catchments where there was a decline in validation performance (A2390512 and A2390519), the variable X2 parameter with the change in catchment averaged groundwater level produced a significant improvement in the model performance on the validation periods.

3.3 Conclusions

Runoff models are useful tools to inform water planning and management. However, trends in the rainfall-runoff relationship may result in a model calibrated to one period producing misleading estimates of future water availability. This study has considered two approaches to representing a change in catchment dynamics (termed non-stationarity) not represented by the rainfall record alone. The first is to bring more information into the model framework to represent the non-stationarity, the second is to modify the model structure in an attempt to improve the ability to represent such changes. In total, seven model configurations have been tested for their ability to simulate flow for periods with different runoff relationships. A number of steps have been taken to increase the robustness of the comparative results: 1) three catchments have been considered, and 2) three calibration algorithms have been used to provide greater confidence that the parameter sets identified are near optimal. The results indicate that for cases when the model performance deteriorates in a validation period compared to a calibration period, varying model parameters using trends in groundwater data improved model performance. For the case of the GR4J model used, varying the X2 parameter was the most promising parameter to improve model performance.

However, a threshold behaviour in the model parameter values that co-varied with the groundwater trend information was observed, where one value was most appropriate under one set of conditions, and a second value after a change in groundwater level of a certain magnitude occurred. This is opposed to a smooth relationship between the model parameter value and the groundwater level. As such, for short term forecasting, calibrating the model parameters to simulate the most recent data, representing the current catchment conditions, may be a suitable approach to determine the model parameters, with the unnecessary complication of introducing further input data requirements and increasing model complexity. However, for longer time simulations, such as for water allocation planning, accounting for trends in the catchment, through changes in groundwater level or otherwise, is likely to be a beneficial approach.

4 Uncertainty Estimation for Conceptual Rainfall Runoff Models

The objective of this section is to address the study objective of developing lumped conceptual rainfall-runoff models to provide predictions with uncertainty of monthly streamflow in Drain M given a seasonal climate forecast. The previous section identified a threshold behaviour in the relationship between groundwater data and model parameters, where the groundwater data were essentially being used to switch between one set of model parameters and another set, after a certain change in the groundwater level had occurred. The previous section also found there was limited benefit in including groundwater dependent parameter values for cases where there was not a distinct change in the rainfall-runoff relationship. For the purposes of short term forecasting, the simpler model with the smallest data requirements is beneficial, particularly as real time groundwater data are not regularly available through telemetered systems, as is the case with climate and streamflow data. As such, in this section the standard form for the GR4J conceptual model has been adopted, and instead only more recent data that is expected to have a more stationary rainfall-runoff relationship, and be more representative of the catchment conditions, is considered.

For the purposes of predicting streamflow, it is desirable to not only provide a value for the prediction, but also quantify the uncertainty associated with the prediction. Many approaches are available to quantify the uncertainty in model predictions, from identifying a range of model parameters that represent the behaviour of the catchment using approaches such as the generalised likelihood uncertainty estimation (GLUE) approach (Beven et al., 2008), to approaches that attempt to characterise each individual source of error explicitly (e.g. Kavetski et al., 2003; Vrugt et al., 2005). In this work, predictive uncertainty has been estimated using an intermediate aggregation approach, where the total uncertainty is characterised by assessing parameter uncertainty explicitly, and aggregating all other sources of uncertainty into a post processed statistical model of the residual error. This residual error model provides a statistical description of the differences between the model predictions and observed data, without trying to disentangle its contributing sources (Evin et al., 2014). This approach is generally less data intensive and less complex than the disaggregation approaches, which attempt to characterise each individual source of error.

The rainfall runoff model has a number of parameters that must be specified to provide streamflow predictions. The value of the model parameters is uncertain for a number of reasons, such as the simplification of the model in representing the rainfall – runoff generating processes in a catchment, and errors in the climate input and streamflow data used to identify the parameter values. In order to assess the suitability of a set of model parameters a likelihood function is required, which provides a quantitative measure of the likelihood that a set of model parameters are representative of the catchment of interest, given the observed data. Different likelihood functions emphasise different components of the flow regime, for example focus on the peaks of the hydrograph when high flows or flooding risk is of interest, or by contrast the long term catchment yield may be of most interest for water planning. The likelihood function selected will determine the parameter values that are adopted as representative of the catchment of interest, for example Cheng et al. (2014) found the form of the likelihood function had a significant impact on the calibrated parameters and the simulated results of high and low flow components.

The parameter uncertainty only represents one component of the total predictive uncertainty, and in this work the remaining component of the predictive uncertainty has been represented using a residual error model. Adequate representations of residual errors necessary for reliable and precise hydrological predictions. If poor assumptions are made about the properties of residual errors, unreliable or highly uncertain predictions are obtained (Thyer et al., 2009). For example, residual errors in hydrological applications are generally heteroscedastic (i.e. the variance in the model errors is not constant, and increases with the magnitude of the flow) and autocorrelated (Sorooshian and Dracup, 1980), which should be accounted for to ensure reliable uncertainty predictions.

4.1 Aims and objectives

The aim of this section it to determine a methodology for estimating the total predictive uncertainty of lumped conceptual rainfall runoff models used to predict the monthly flow volume one month in advance in the catchments contributing to Drain M.

A number of objectives have been identified in order to meet this aim. These were to test:

1. When estimating total predictive uncertainty using residual error models, what strategies can be adopted to improve the reliability and precision of the prediction distribution? In particular:
    a. Methods to account for the heteroscedastic nature of the model errors
    b. Methods to correct for bias in model errors
2. When estimating parameter uncertainty, do different likelihood functions alter the reliability and precision of the model predictions?

The methodology used to address these objectives are outlined in the following section.

4.2 Methodology

The lumped conceptual rainfall runoff model selected was GR4J, as outlined in Section 3. The rainfall runoff model has a number of parameters that must be specified to provide streamflow predictions. The residual error models also have parameters that must be identified for a given assumed form. Residual error model parameters can be estimated in a combined process with the runoff model parameters, or in a postprocessor step after the parameter estimation process. Limitations of the postprocessor approach include it must be undertaken for each set of runoff model parameters separately, and ignores interactions between the runoff and residual error model parameters. However, Evin et al. (2014) found the postprocessor approach can be more robust than inferring all model and residual error parameters concurrently, because of the interactions between the parameters of the different models, which mean it is difficult to identify representative values, resulting in unrealistically wide uncertainty limits. As such, the postprocessor approach has been adopted for this work. The approaches used to identify parameter uncertainty, total predictive uncertainty using residual error models, and metrics used to assess model performance are outlined in this section.

4.2.1 Estimation of Parameter Uncertainty

Posterior parameter uncertainty has been estimated using the DiffeRential Evolution Adaptive Metropolis (DREAM) approach (Vrugt et al., 2009). The Hydromad R package (Andrews et al., 2011) implementation of the DREAM algorithm and GR4J model have been adopted.

4.2.1.1 Prior Parameter Distributions

Uniform prior distributions have been adopted for all parameters, with the assumed bounds given in Table 5. A number of trial runs were used to ensure the bounds of the prior distribution were suitable, and resulted in limited density of the posterior distribution occurring at the bounds specified after inference of the parameter distributions.

Table 5 Bounds adopted for the uniform prior distribution

Parameter	Lower Bound	Upper Bound
x_1	100	600
x_2	-15	5
x_3	1	300
x_4	0.5	6
split	0.6	0.99

4.2.1.2 Data length for Parameter Estimation

Typically, rainfall runoff models are developed using as long a period as possible (subject to retaining an independent validation period), as it is generally considered that the longer the calibration period, the more robust the parameter set identified will be (Brigode et al., 2013). This approach assumes that the parameter values derived from the calibration period are valid for the prediction period (Vaze et al., 2010). This assumption has been questioned recently because of possible changes in the stationarity of hydrologic processes in some catchments (Luo et al., 2012).

To address the issue of changes in the rainfall-runoff relationship identified in the previous section, the approach of Luo et al. (2012) to minimize the impact of external influences on the model parameters inferred has been adopted.

External influences include model structural limitations, physical changes in the catchments influencing the rainfall runoff relationships (land use change, change in groundwater level), or over-representation of particular hydroclimatic conditions. In this approach, the model parameters were recalibrated each year in a rolling approach based on the data from the preceding period. The parameter values identified were used to simulate the following one year of data (following that used in the calibration process), before being recalibrated again. The recalibration methodology adopted allows for the change in parameter distributions over time to be adopted.

A 10 year calibration period has been adopted as a trade-off between a longer calibration period expected to reduce the parameter uncertainty, and a shorter calibration period representing the catchment conditions most representative of the catchment when used for the one year forecast period. As an example, a 10 year calibration period may be 1/5/1995-30/4/2005, after a one year warmup period. The prediction period was then considered to be the following one year (1/5/2005-30/4/2006). The process was then repeated, where in this example the calibration period becomes 1/5/1996-30/4/2006 and the prediction period 1/5/2006-30/4/2007. Through this rolling approach, the model predictions assessed were independent of the data used for calibration in all the results presented.

4.2.1.3 Likelihood Function

Once a representative period of data is determined, the suitability of a set of model parameter values in representing a period of observed flow data is typically assessed using a pre-specified likelihood function. In an uncertainty estimation framework, the likelihood function represents an assessment of how well the parameters represents the catchment. This is analogous to an objective function calculating the error between observed and simulated flows in a traditional optimisation procedure. Different likelihood functions emphasise different components of the flow regime, for example focus on the peaks of the hydrograph when high flows or flooding risk is of interest, or by contrast the long term catchment yield may be of most interest for water planning. The likelihood function selected will determine the parameter values that are adopted as representative of the catchment of interest, for example Cheng et al. (2014) found the form of the likelihood function had a significant impact on the calibrated parameters and the simulated results of high and low flow components.

Two different likelihood functions have been tested: sum of squared errors, and a combined likelihood function. A sum of squared errors (SSE) likelihood function is derived from an error model assuming independent, homoscedastic residuals. While this is unlikely to be the case for hydrological applications, statistics based on sum of squared errors (Nash Sutcliffe, R^2, RMSE) are often used for model calibration and assessment, and hence this function has been selected. This function has also been adopted given the focus of this study is on the estimation of monthly runoff volumes, as this function provides a focus on the highest flows in the time series, where the majority of the runoff volume occurs.

The second likelihood function considered adopted a combined multi-objective approach (CMB) (Gupta et al., 2009). The terms of the function adopted are based on those suggested for surface water calibration using the Model-Independent Parameter Estimation and Uncertainty Analysis tool, PEST (White et al., 2014), and also including the total volume bias term of Viney et al. (2009):

$$E = R^2\left(Q_{d,sq}\right) + R^2(Q_m) + R^2(Q_{FDC}) - 5|\ln{(1 + B)}|^{2.5} \tag{6}$$

where $Q_{d,sq}$ is the square root transform of the daily flows, Q_m are the monthly volumes, Q_{FDC} is the flow duration curve, calculated as the flow exceeded at five percentile intervals, and B is the total volume bias as a percentage. The R^2 statistic is used to balance the three objectives by normalising by the observed variance, which will all tend toward one as the metric improves, and the total volume bias used as a penalty term against higher values. Adopting different time periods for the different components reduces the mathematical relationship between the components, which should be considered to ensure that the combined function is balanced (Węglarczyk, 1998). This can be considered an informal likelihood function, as it is not derived from an assumed error model.

4.2.1.4 DREAM Configuration

A maximum of 25,000 evaluations were used, and 25% of these evaluations were used for a burn in period, where the initial samples were discarded before the Markov Chain was assumed to have stabilised. The number of parallel chains was set to the number of parameters, i.e five parallel chains were adopted (Vrugt et al., 2009). Autocorrelation between sequential samples was used to thin the results and identify an efficient sample of parameter sets that represented the posterior parameter distribution.

Before using the parameter uncertainty (represented by the posterior parameter distribution) to calibrate residual error models, it is important to ensure that the DREAM uncertainty estimation method is performing adequately. A number of diagnostics suggest good mixing and convergence of the DREAM algorithm has been achieved. Traceplots of the iteration number against the value of the draw for each parameter indicated good mixing across all parallel chains. This is confirmed by plots of the posterior marginal distributions, where the parameters do not contain substantial density near the assumed limits of the uniform prior distribution. However, some additional iterations were undertaken to ensure this was the case, which informed the bounds adopted for the prior distributions (Table 5). Median acceptance rates were the in range of 30-40%, also suggesting that the prior distributions were not too wide (which would result in low acceptance rates) or too narrow resulting in poor mixing between the chains (indicated by a high acceptance rate).

Good mixing is also demonstrated by the autocorrelation between draws of the MCMC sampler, which decreased over increasing lags, with insignificant correlation after 20-30 draws. As noted above, this autocorrelation was used to thin the MCMC results to remove autocorrelation and achieve an efficient sample of the posterior parameter distributions.

The Gelman and Rubin's shrink factor, which measures the difference between the variance within and between the chains, can also be used to indicate that adequate convergence has occurred. For the analyses undertaken, the 97.5th percentile upper bound value started at values in the range of R = 1.5 - 2.5, and converged to R = 1.01 - 1.02 within the evaluations undertaken. By comparison, Vrugt et al. (2009) suggested R < 1.2 could be considered to represent convergence. Based on these diagnostics, it has been concluded the DREAM algorithm provided an accurate estimate of posterior parameter distributions, suitable for further analysis of the predictive uncertainty.

4.2.2 Estimation of Predictive Uncertainty

The parameter uncertainty was estimated at the daily time scale, given this is the typical scale for simulating streamflow for yield hydrology, and it is assumed that if the model represents the daily flows well, then it will be more robust when the results are presented as a monthly volume. However, the residual error model was fitted to the errors in the monthly volumes, as this is the main time scale of interest for the management of the drains in the South East. While fitting the residual error model to monthly data reduces the number of data points to estimate the model parameters, the aggregation to monthly reduces much of the autocorrelation in error between observed and simulated monthly volume from one time step to the next (Evin et al., 2014), allowing simpler error models to be considered while still providing robust results.

This aggregation removes one challenge with fitting models of residual errors, where another is the variance in the errors is not constant, and tends to increase with the magnitude of the simulated flow. Typically, residual error models assume some form of statistical distribution with a mean of zero. However, the modelled flow values may be biased in one direction or another, more often over or underestimating the observed flow on average. Two approaches for considering the non-constant variance (heteroscedasticity) and bias in modelled errors have been tested to select the residual error model to adopt for the Drain M catchments.

4.2.2.1 Methods to Account for Heteroscedasticity

A variety of strategies have been proposed to address this heteroscedasticity in errors. For example, heteroscedasticity is often taken into account by transforming observed and predicted streamflows using logarithmic or Box-Cox transformations before applying constant-variance statistical error models (Kuczera, 1983; Morawietz et al., 2011). This approach represent an indirect treatment of heteroscedasticity, by transform distribution of errors to become more Gaussian, and hence can be represented by a random normal variate, N(0,σ2). Another approach that has been adopted is to represent the heteroscedasticity explicitly, where the variance of the residual errors is a function of the simulated runoff magnitude. The direct conditioning strategy has been used in a number of studies

(Schoups and Vrugt, 2010; Sorooshian and Dracup, 1980; Thyer et al., 2009). Both direct and transformed approach have been considered in this work to account for the heteroscedasticity in model errors.

A residual error model based on additive errors, ε, has been assumed at a monthly timestep t, as:

$$\varepsilon_t = \widetilde{Q_t} - Q_t^\theta \tag{7}$$

Where $\widetilde{Q_t}$ is the observed monthly flow at time step t, and Q_t^θ the corresponding simulated monthly flow based on model parameter set θ. To derive a model of this error, both direct and transformation approaches to account for the heteroscedasticity in errors have been tested:

- Direct: ε_t is assumed to be N(0,σ), with $\sigma_t=\alpha_0 + \alpha_1 Q_t^\theta$, to allow the variance of the errors to increases with the simulated flow (Evin et al., 2014). In this case, the error model parameters are α_0 and α_1.
- Transformed: ε_t is assumed to be N(0,σ_1), but first a logarithmic transform is applied, i.e. $\varepsilon_t = \log(\widetilde{Q_t} + \alpha_0) - \log(Q_t^\theta + \alpha_0)$. The constant α_0 is used to allow for the long transform of a zero monthly flow to be computed. The error model is assumed to have constant variance after the transformation, σ_1. As with the first case, the error model has two parameters, α_0 and α_1.

The residual error model parameters were estimated for each parameter set sampled from the posterior parameter distribution. A maximum likelihood approach was used, accounting for the Jacobian of the transformation for the transformation approach. A sample of 100 parameter sets were taken from the posterior distributions identified by the DREAM algorithm, and for each a further 100 replicates from the fitted residual error model were derived, resulting in 10,000 samples of the total predictive uncertainty. For the direct method for accounting for heteroscedasticity the terms are all additive. For the transformed approach, the total error flow $Q_t^{\theta,\alpha}$ can be substituted for $\widetilde{Q_t}$ as:

$$\varepsilon_t = \log(Q_t^{\theta,\alpha} + \alpha_0) - \log(Q_t^\theta + \alpha_0)$$

$$Q_t^{\theta,\alpha} = e^{\varepsilon_t}\left(Q_t^\theta + \alpha_0\right) - \alpha_0 \tag{8}$$

4.2.2.2 Bias Correction to Simulated Flows

The residual error models considered assume normally distributed errors (after a log transformation in the transformation case) with a mean error of zero. However, often this is not the case, Evin et al. (2014) states that as the postprocessor method use an "incomplete" error model when calibrating the hydrological parameters (e.g., the error autocorrelation) the parameter estimates obtained can be unreliable (biased and/or with poor uncertainty estimates). This becomes even more difficult for ephemeral catchments with complex error structures, and also cannot be tested when informal likelihood functions, such as the CMB function, are adopted.

In the absence of an accurate error model to form the likelihood function for parameter estimation, a number of strategies have been considered to account for any bias in the simulated flows:

- Bias Corrected: A constant was subtracted from the simulated flows for each parameter equal to $E(\widetilde{Q_t}) - E(Q_t^\theta)$. The transformed values of Q_t^θ were then used to calibrate the error model as above. This transformation can introduce negative values for Q_t^θ, which were set to zero prior to estimating error model parameters.
- Assimilated: The simulated flow was corrected to match the observed flow at the start of each month by modifying the level in the routing store. This approach assumes that the forecast is undertaken at the start of each month, and the observed flow is known. This approach is outlined in more detail below.
- Assimilated and Bias correction: Both assimilation of observed flows into the model routing store during the model run, and bias correction of the simulated flow after the model run.

The approach of Demirel et al. (2013) has been adopted to incorporate the current observed flow into the model states at the start of each month. The model was run for a one year warmup period, and the simulated level in the production store at the end of this period was maintained. As model states generally do not reflect reality directly (Berthet et al., 2009) the production store level has been maintained "as seen" by the model.

In order to force the model to simulate the flow observed at the start of the month, the necessary routing store level to produce the observed flow after accounting for the modelled direct flow was calculated (Demirel et al., 2013). In GR4J, total streamflow, Q_t^θ is calculated as the sum of the flow direct from the production store (after applying a unit hydrograph), $Q_{t,d}^\theta$, and the flow from the routing store, $Q_{t,r}^\theta$. The modelled $Q_{t,d}^\theta$ is subtracted from the observed flow, $\widetilde{Q_t}$, to derive $Q_{t,r}^\theta$ necessary for $Q_t^\theta = \widetilde{Q_t}$. The routing store level, R, can then be solved for using the equation used by the GR4J model to calculate the outflow from this storage:

$$Q_{t,r}^{\theta} = R\left(1 - \left(1 + \left(\frac{R}{x_3}\right)^4\right)^{-1/4}\right) \tag{9}$$

where x_3 is an estimated runoff model parameter. In the case where $Q_d > Q_o$, R was set to zero. Assimilation of the routing store was applied separately to each parameter set sampled from the posterior distribution.

4.2.3 Assessment Metrics

In order to assess the impact of the different likelihood functions and calibration periods on predictive uncertainty, two metrics that quantify predictive uncertainty have been used. These metrics have been selected to represent two different aspects of predictive uncertainty: (1), including how reliable the distribution is; and (2) the resolution, or precision, of the distribution.

Predictive uncertainty has been calculated from the model parameter uncertainty identified for each case, by calculating the flow exceeded at five percentile intervals. The resulting predictive distribution (PD) has been assessed using the approach and terminology of Renard et al. (2010), namely the “reliability” to quantify the statistical consistency between the observed data and the PD, and “resolution” or precision to quantify the sharpness of the PD.

In time series analysis, evaluating a PD requires comparing a time-varying random variable X_t (with cumulative distribution function F_t) to a time series of realizations x_t, where in this case x_t represents the observed runoff, and X_t the range in streamflow simulated from the parameter uncertainty. If the PD is reliably quantified, the observations should correspond to realizations from the PD. If the realizations X_t are consistent with F_t, the p values $F_t(x_t) = p(X_t \leq x_t)$ will follow a uniform distribution on the interval [0,1]. Deviation from the uniform distribution indicates over or underestimated uncertainty, which has been assessed using Renard et al. (2010):

$$\alpha = 1 - 2\frac{\sum_{i=1}^{N}\left|p_{x,i} - p_{th,i}\right|}{N} \tag{10}$$

Where $p_{x,i}$ and $p_{th,i}$ are the i^{th} observed and theoretical p values of x_t, N is the number of x_t values. The value of α varies between $\alpha=0$ representing the worst reliability, occurring when all observed p values equal either 0 (whole PD exceeded the observed value) or 1 (whole PD below the observed value), and $\alpha=1$ (perfect reliability).

“Resolution” denotes the sharpness or precision of the PD, as two results can both yield the same α value in terms of the reliability of the PD, but with different resolutions, or widths of the uncertainty distribution. The relative resolution metric of Renard et al. (2010) has been used to assess the resolution of the PD, calculated as:

$$\pi^{(rel)} = \frac{1}{N}\sum_{t=1}^{N}\frac{E[X_t]}{Sdev[X_t]} \tag{11}$$

Where E[] and Sdev[] are the expectation and standard deviation operators. Low values of $\pi^{(rel)}$ indicate wide uncertainty distributions, where the standard deviation is greater than the average, where high values indicate narrow distributions, with a low standard deviation relative to the average value.

Months a simulated volume less than 0.1 ML/month have been ignored in the assessment of the reliability and resolution of the PD, as it is difficult to assess model reliability when all of the density of the PD is at zero, and the observed value is also zero. This represents an accurate and reliable estimate of the flow, but the p value for this occurrence is difficult to calculate, as p values of 0 or 1 indicate poor reliability. Also, very low flows typically have relatively large errors that almost arbitrarily influence the assessment of the PD. For example, in a study of 36 gauges in Australia, Tomkins (2014) found most gauges showed significant deviations at low stages, affecting the determination of low flows.

4.3 Results

The results are presented in three sections to address the three different objectives of this section. These were, for the South East catchments considered, what was the best:

1. Approach to account for the heteroscedasticity in errors,
2. Approach to account for the bias in errors,
3. Likelihood function to adopt to reliability estimate monthly runoff volume

4.3.1 Strategies to account for Heteroscedasticity in errors for South East catchments

The reliability and precision of the predictive distribution (combined parameter uncertainty and residual error uncertainty), PD, for the two likelihood functions and two methods for accounting for heteroscedasticity in the model errors is presented in Figure 20. It can be seen in all cases the log transformed method (purple shapes) resulted in more precise PD than the direct approach (i.e. has a higher value on the y-axis compared to the same shape in orange).

For catchments A2390512 and A2390519, the log transformed approach also resulted in a more reliable PD than the direct approach (i.e. for the same shape, the purple shape has a higher value on both axis compared to the orange shape). For catchment A2390514 the transformed approach still improved the precision of the PD, however in this case there was a trade off with a slight reduction in the reliability of the PD.

Often there is often a trade-off between reliability and precision, the PD can be made to be more precise and capture a smaller range of flows, but this typically comes at the expense of the reliability of the predictions. Given that the log transform approach improved both metrics for two catchments, and the precision for all three catchments, the log transform has been selected as the best approach to represent the residual error for the catchments considered.

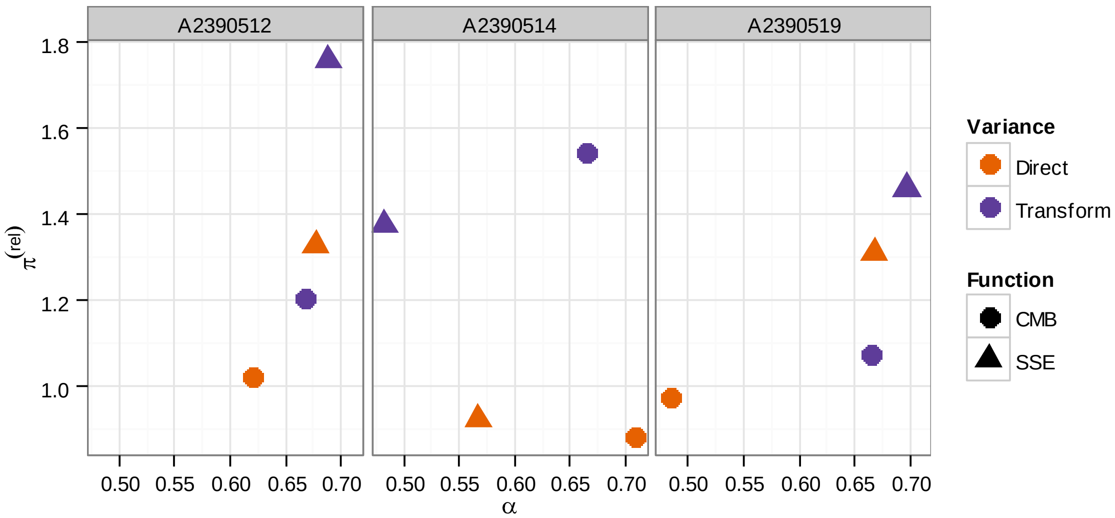


Figure 20 Reliability (α) and precision ($\pi^{(rel)}$) for direct and transformation methods for accounting for heteroscedacity in model errors

4.3.2 Strategies to correct Biases in errors for South East Catchments

The effect on the reliability and precision of resulting PD by correcting any biases in the modelled flows compared to observed, to better match the assumptions of the residual error model, are considered in this section. The change in reliability and precision compared to not correcting biases at all is presented in Figure 21, where the origin represent the value of no bias correction, and positive values indicate an improvement by the correction method.

While the simple bias correction improved the precision of the PD for catchments A2390512 and A2390519, this approach had a limited effect on the reliability for these catchments. For catchment A2390514, the bias correction approach had a variable impact on the PD depending on the likelihood function adopted.

The assimilation of observed flow into the routing store level improved reliability in all cases, however the effect on precision was mixed (limited change for catchment A2390519, some improvement for catchment A2390512 and reduced precision for catchment A2390514). Using both approaches dramatically reduced the improvement in reliability gained by the assimilation approach for catchment A2390512, maintained a similar result for catchment A2390519 and reduced the reliability for catchment A2390514.

Given the improvement in reliability in all cases by the assimilation only approach, assimilation of the observed flow into the routing store has been adopted as the bias correction approach in this work.

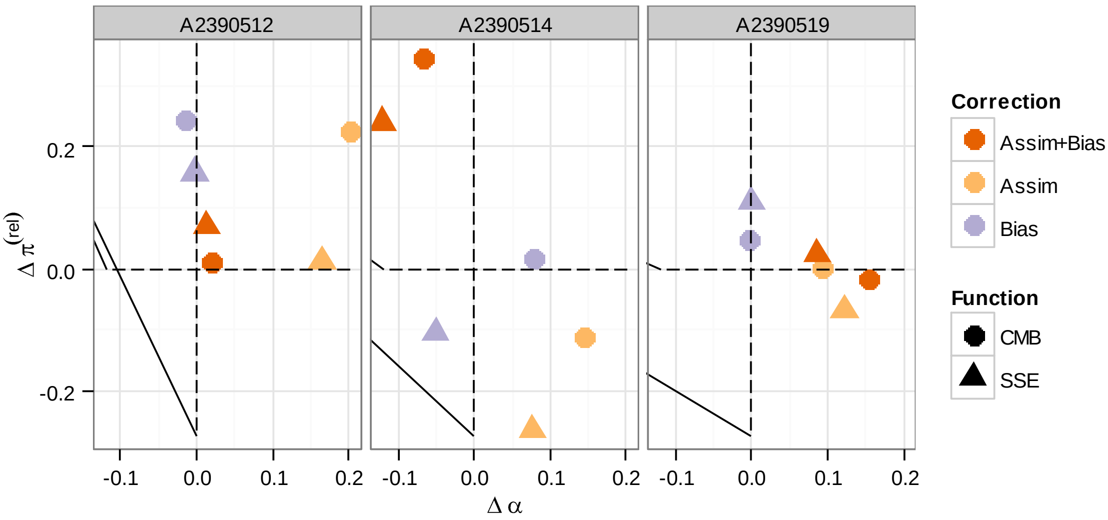


Figure 21 Change in reliability (α) and precision ($\pi^{(rel)}$) due to different bias correction methods

4.3.3 Likelihood Functions to assess Parameter Uncertainty

The same information presented in Figure 21 is also presented in Figure 22, this time as the absolute values rather than the change compared to no bias correction. When considering the assimilation only correction approach (light orange shapes), it can be seen that the SSE likelihood function resulted in improved reliability and precision for catchment A2390512 compared to the CMB likelihood function, and a similar reliability and improved prevision for catchment A2390519. For catchment A2390514 the precision was similar between the two functions, however the CMB likelihood function produced the highest reliability.

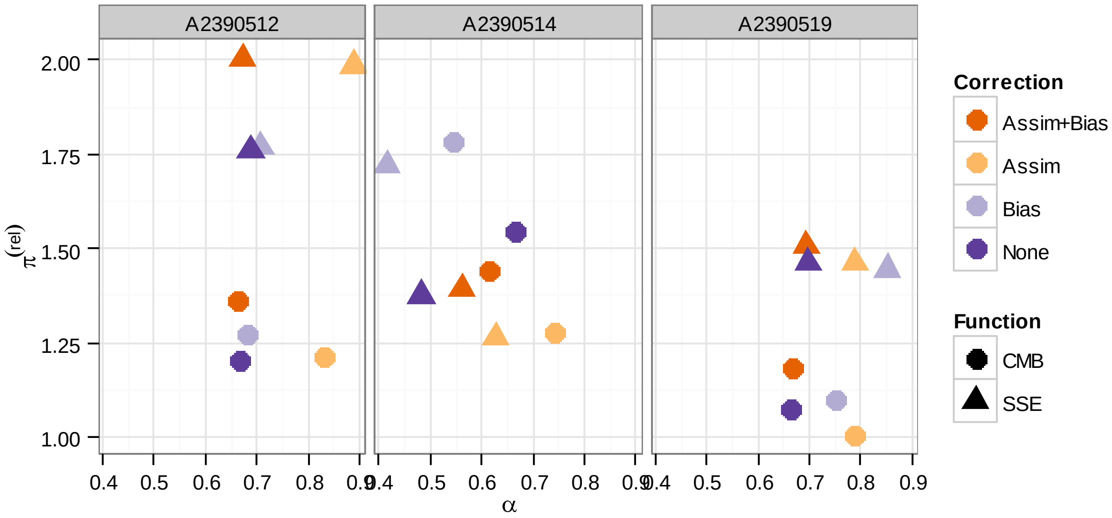


Figure 22 Reliability (α) and precision ($\pi^{(rel)}$) for the different likelihood functions and bias correction methods

Time series of the median monthly model prediction, with 95th percentile confidence bounds, compared to the observed monthly volume is presented in Figure 23 for the SSE likelihood function, and Figure 24 for the CMB likelihood function. The improvement in precision, as the reduced width of the uncertainty bounds, can be seen for catchments A2390512 and A239019. The difference in reliability for catchment A2390514 cannot be seen from the uncertainty bounds alone, however the SSE likelihood function tends to estimate the flow events in 2003 and 2004 slightly better for this catchment (both median and uncertainty bounds). As such, the SSE likelihood function has been adopted to estimate parameter uncertainty for all three catchments.

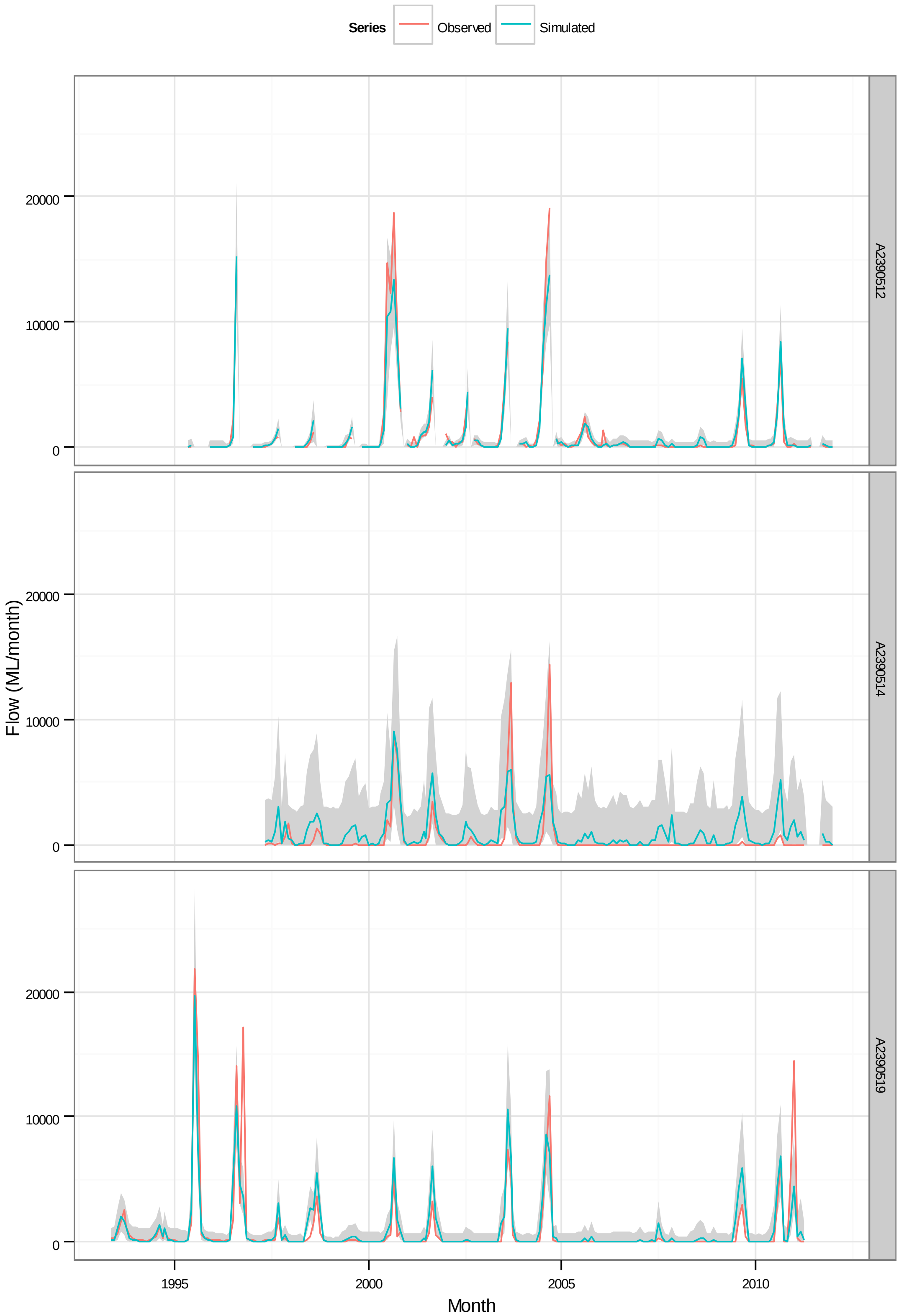


Figure 23 Median simulated flow with 95th percentile confidence bounds compared to observed for the SSE likelihood function

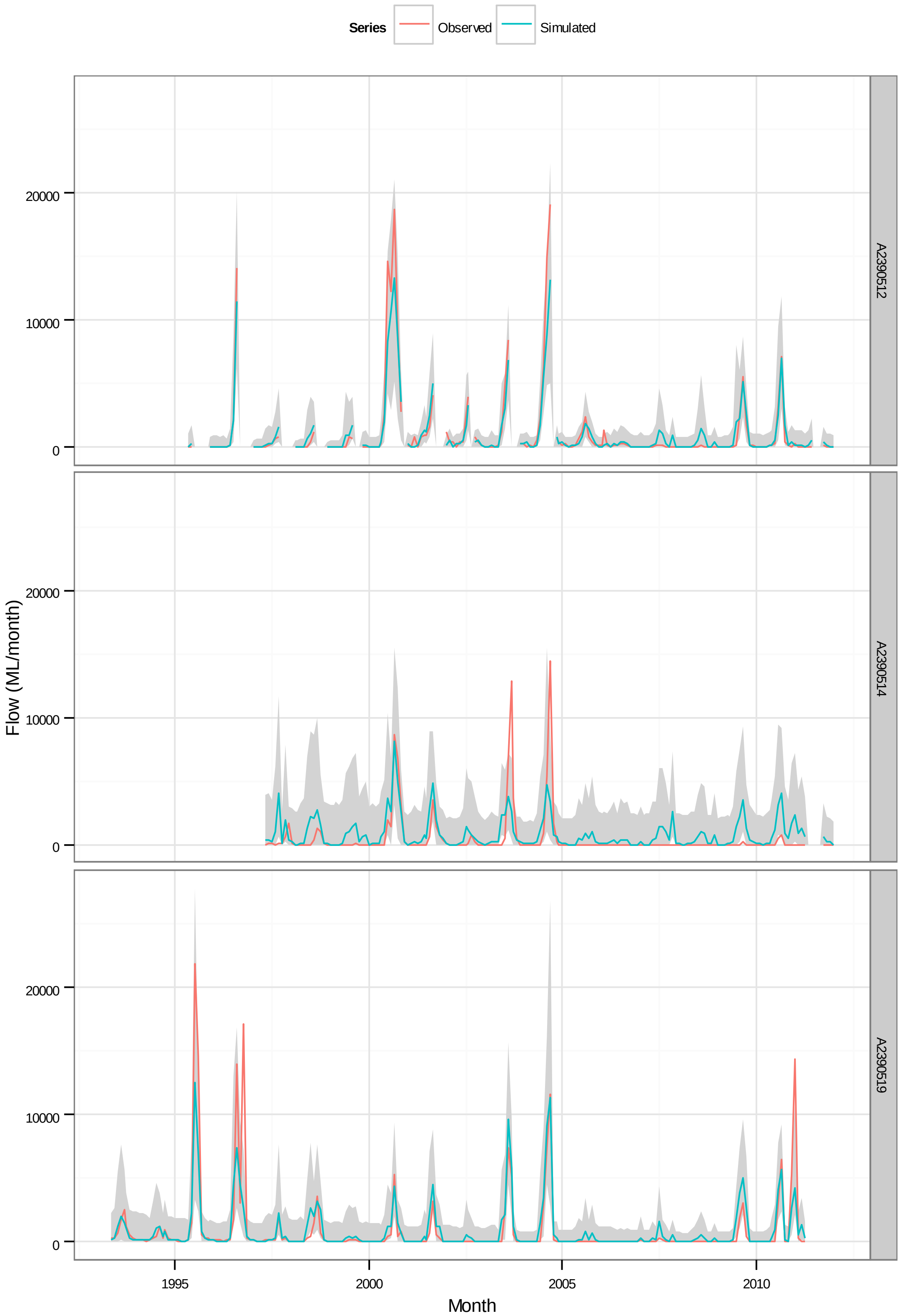


Figure 24 Median simulated flow with 95th percentile confidence bounds compared to observed for the CMB likelihood function

4.4 Discussion and Conclusions

For the purposes of predicting streamflow, it is desirable to not only provide a value for the prediction, but also quantify the uncertainty associated with the prediction. For this work an aggregated approach was adopted, where the total predictive uncertainty was estimated, but the contributions of different sources (e.g input data, parameter, output data, model structure) to the total predictive uncertainty were not quantified. Within this aggregated approach, a postprocessor residual model error approach was adopted, as this allows the model parameter uncertainty and residual error model parameters to be identified separately, and can lead to more robust estimates of the model uncertainty compared to jointly assessing all parameters (Evin et al., 2014).

For the catchments considered, the transformed approach for incorporating the heteroscedasticity in model errors was found to provide more precise and in most cases also more reliable predictive distributions than the direct approach. This decrease in precision without a corresponding increase in reliability for the direct method implies that the variance was overestimated by direct approach, and may be able to be improved by considering different functional forms for the relationship between simulated flow and the variance in errors.

Three approaches to address the bias in model estimates were considered. A simple correction by subtracting a constant from the simulated values improved the precision of the predictive distribution, but in most cases did not improve the reliability. Assimilation observed flow data to correct the routing store of the model was found to improve the reliability of the estimates in all cases considered. This approach effectively corrected the bias in the modelled flow on a monthly basis within the model, and hence allows for different corrections each month. The assimilation approach also allows more data to be included in the model which can lead to improved predictions. While changes to the model storage levels during a simulation could be considered to compromise the overall water balance by changing the model states, this approach is likely to be more coherent that the simple bias correction. Somewhat surprisingly, the combination of both assimilation and bias correction did not further improve the predictive distribution compared to the assimilation only approach. This may be due to the simplistic subtraction of a constant, and a time dependent correction, a constant for each season or month for example, or flow dependent correction, by quantile- quantile mapping, may improve the performance of the bias correction approach.

Two different likelihood functions were considered to identify the model parameter uncertainty. The different likelihood functions were found to affect which parts of the flow regime the model represented more accurately, with the SSE function found to estimate the peaks in the hydrograph better, where the majority of the flow volume occurs. The predictive uncertainty was also greater for the CMB function, possibly because this more complex function considered multiple aspects of the hydrological regime, making it more difficult to identify representative sets of model parameters.

This section has developed a method to simulate one month ahead flow volumes for the catchments contributing to Drain M using lumped conceptual rainfall runoff models. A number of approaches to account for the uncertainty in the predictions were tested, allowing the most reliable and precise quantification of the predictive uncertainty to be selected from the approaches considered. However, the lumped conceptual rainfall runoff models are only one method to predict streamflow, and in the following section data driven approaches are developed.

5 Incorporating Soil Moisture into Statistical Flow Forecasting Models

As mentioned in Section 1, an objective of this study was to develop both lumped conceptual and statistical models to provide forecasts of upcoming streamflow in Drain M, such that the suitability of both of these modelling approaches, as well as a hybrid approach, could be assessed and compared in the ephemeral catchments of the South East. This section, together with Section 6, presents the development of the statistical models, as well as hybrid models, where outputs from the lumped conceptual model are used as inputs to the statistical model.

The models developed in this study should help to address the management question of how best to maximise the use of water in Drain M in order to support the requirements of Lake George at the termination of the drain, as well as the more northern wetlands in the USE. As such, the main focus in this section is on the development of accurate and reliable statistical and hybrid models for forecasting flow one month in advance at site A2390512 (point C in Figure 3), since forecasts at this location provide important information about volumes flowing into Lake George, which is required before questions regarding any diversions to the north can be addressed. Details of the model development steps are presented, with particular emphasis on the selection of optimal model inputs and the inclusion of different sources of data for best representing initial catchment conditions. In Section 6, the statistical and hybrid models are then compared to the lumped conceptual models across all three subcatchments.

5.1 Statistical flow forecasting

The GR4J CRR model outlined in Sections 3 and 4 derives the streamflow response of a catchment through the simplified representation of key components of the hydrological system. Statistical or data-driven flow forecasting models, on the other hand, characterise system response primarily through the extraction of information contained in a set of observed data, without explicit consideration of the physical processes occurring within the hydrological system (Kokkonen and Jakeman, 2001; Toth and Brath, 2002). Consequently, these models are not limited by an inadequate or unsuitable description of the underlying physical processes as CRR models may be (e.g. unlike the GR4J model, changes in streamflow due to changes in groundwater level may be characterised within a statistical forecasting model given adequate predictor information). However, the performance of statistical flow forecasting models is highly dependent on the availability and quality of observed data. Ideally, concurrent observations of all relevant predictors and the streamflow response would be required, with records sufficiently long to include a wide range of conditions; while in reality, statistical models usually need to make do with whatever data are available. On the other hand, statistical models do not require specific input information as conceptual models do, but rather they are able to utilise whatever data are available to their best advantage. However, with no mathematical descriptions of the underlying physical processes, such models typically have limited or no extrapolation capability, meaning that they are only valid over the range of the data used for model calibration (however, they can be updated as new data become available, see Bowden et al. (2012) for example).

In order to provide forecasts of future streamflows, statistical flow forecasting models require predictors that describe the initial catchment conditions and the effect of climate during the forecast period. Climate indices based on sea surface temperatures and atmospheric pressure (e.g. Niño-3, Southern Oscillation Index) tend to be used as predictors of future climate, while observations of antecedent streamflows and rainfall are typically used to represent the initial catchment condition (Robertson et al., 2013). In temperate regions such as the Lower South East, the states of the soil moisture and groundwater stores are the most relevant for seasonal streamflow forecasting (Robertson and Wang, 2012). Yet, the use of antecedent rainfall and streamflows as proxies for initial catchment conditions often provides a rather crude approximation of initial catchment wetness. For example, while soil moisture and groundwater storages have a finite capacity, rainfall and streamflow can continue to increase beyond levels that represent the status of these stores and, as a result, high antecedent streamflow and rainfall data may result in streamflow forecasts that are too high. On the other hand, low antecedent streamflows may lead to the underprediction of upcoming streamflows if the catchment is wetting up and antecedent rainfall is replenishing soil and groundwater storages rather than converting to streamflow. This is particularly pertinent in low-yielding ephemeral catchments, where periods of zero flow provide no information on the status of the catchment's soil moisture stores (Wooldridge et al., 2003).

While a number of studies have demonstrated the benefits of employing observations of soil moisture content in streamflow models (see Anctil et al. (2008); Casper et al. (2007); Gautam et al. (2000); Koren et al. (2008); Tayfur et al. (2014); Wooldridge et al. (2003), for example), the limited availability of such data has prohibited the widespread use of soil moisture measurements for representing initial catchment wetness in statistical models. In-situ soil moisture measurement is expensive and labour intensive, making it impractical to implement the necessary measurement networks on a widespread basis (Anctil et al., 2008; Berg and Mulroy, 2006). In the absence of ground based soil moisture observations, simulated data representing the effects of soil moisture in physically-based or conceptual hydrological models may be used as predictors in statistical streamflow forecasting models (e.g. see Anctil et al. (2004), Robertson et al. (2013) and Linares-Rodriguez et al. (2015)); thus, exploiting the benefits of both physically-based and statistical modelling approaches in a *hybrid* forecasting system. However, the accuracy of such soil moisture information is dependent upon the accuracy of the underlying physically-based or conceptual model and may itself be limited by an inadequate or overly simplified description of catchment processes. Furthermore, the availability of such data requires the development of a physically-based or conceptual model of the catchment under consideration, which, in turn, requires modelling expertise and data that may or may not be immediately available.

Recently, with the emergence of new global land-surface data sets, including satellite-based and reanalysis products, alternative soil moisture data are now readily available for representing initial catchment wetness in statistical flow forecasting models. However, these data tend to be available only at a fairly coarse spatial resolution and their utility in statistical models remains unclear. Several studies have shown there to be some potential for using satellite-derived remotely sensed soil moisture observations to improve rainfall-runoff modelling (seeBrocca et al. (2012); Scipal et al. (2005); and Xu et al. (2014)); however, these have typically involved the assimilation of satellite-derived soil moisture into conceptual and physically-based rainfall-runoff models, while very few studies have utilised such data as predictors in statistical streamflow forecasting models (Bindlish et al., 2009).

5.2 Aims and objectives

In this section, the value of three different sources of soil moisture data for improving one-month-ahead statistical streamflow forecasts is explored. The first source is the modelled water level in the GR4J conceptual soil reservoir, which may be considered analogous to soil moisture, generated using the models presented in Section 4. As such, the models developed using this input information are considered a hybrid of the conceptual and statistical forecasting approaches employed in this study. The second source of soil moisture data was developed as part of the Soil Moisture Climate Change Initiative project, in which soil moisture is recognised to be an “Essential Climate Variable” (ECV) required for characterising the state of the global climate system and enabling long-term climate monitoring. This recently available global record of remotely sensed soil moisture was generated by merging different satellite soil moisture datasets which have been observed since 1978 (De Jeu et al., 2012). Finally, the third soil moisture dataset is derived from a reanalysis using the Predictive Ocean Atmosphere Model for Australia (POAMA), a dynamic climate forecasting system based on a coupled ocean/atmosphere model and ocean/atmosphere/land observation assimilation systems. The models which utilise the POAMA soil moisture data as predictors may also be considered a type of hybrid modelling approach, since these models combine dynamical and statistical forecasts; however, any references to a hybrid modelling approach in this report refer to the combined lumped conceptual and statistical flow forecasting models.

Each of the soil moisture datasets considered provide different information about soil water storage and have their own advantages and limitations. The GR4J derived data provide estimates of initial catchment wetness at the catchment scale based on a model has been developed specifically for the catchment under consideration using local data. However, as mentioned previously, the accuracy of these data is dependent upon the accuracy of the underlying GR4J conceptual model and its suitability for the study region. Within this model, the representation of soil moisture is at a lumped conceptual level, and may not directly represent the physical processes within the catchment (Berthet et al., 2009). As seen in Section 4, this model had varying performance across the three different subcatchments considered in this study. The satellite-derived soil moisture data only provide estimates of soil moisture in the top few centimetres of the soil and at a fairly coarse spatial resolution (~25km). However, being observations, the quality of these data is not dependent on any underlying model; although, they may be biased by the retrieval algorithm used (in regions of dense vegetation, the retrieval algorithm may fail altogether (Dorigo et al., 2012)). Validation of these data over Australia has shown strong correspondence between the monthly mean satellite derived soil moisture data and monthly precipitation and that, temporally, the satellite data provide a good representation of in-situ soil moisture data (Draper et al., 2009). The POAMA soil moisture data, unlike the previous two soil moisture datasets, provide a forecast of average soil moisture over the forecasting period, which may be important when forecasting streamflows at a monthly timescale. However, these data are provided at a much coarser spatial resolution than the satellite data (~200km), as shown in Figure 25, and, like the GR4J data, their accuracy is heavily dependent on the underlying forecasting model. Nevertheless, the POAMA forecast system is currently used by the Australian Bureau of Meteorology (BoM) for providing long range climate forecasts and has undergone extensive validation.

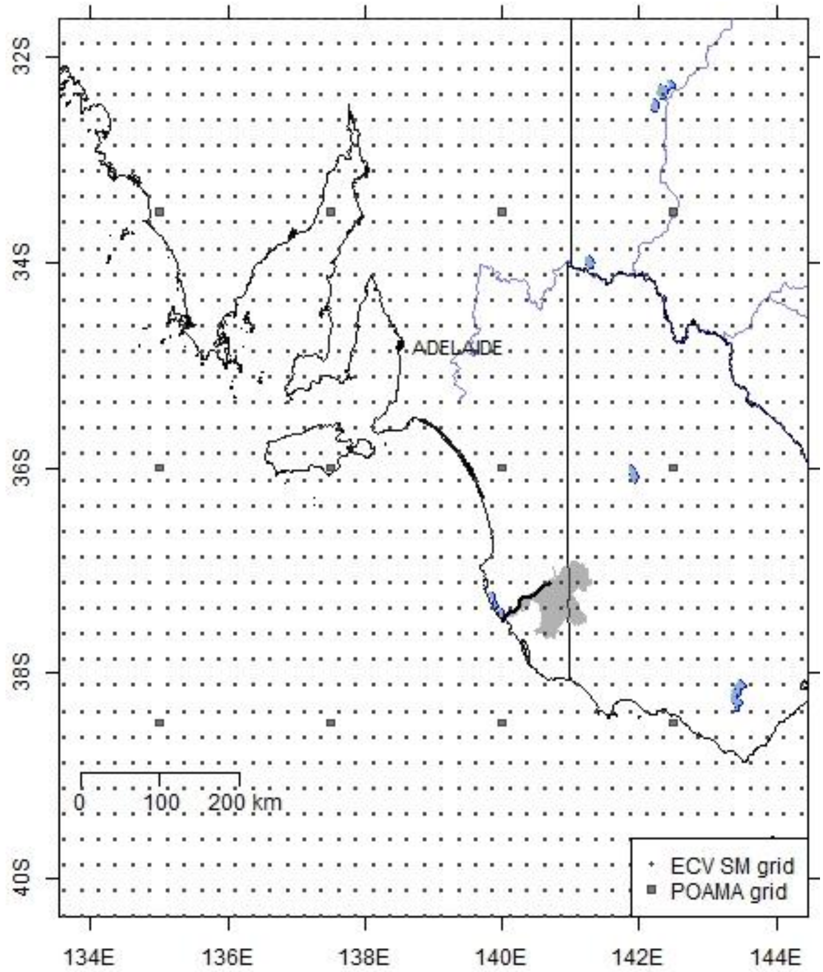


Figure 25 Satellite derived soil moisture and POAMA-2 grid points.

The overall aim in this section is to assess whether the information content of any of these soil moisture datasets, or a combination of them, is sufficient to improve statistical streamflow forecasts, particularly at a monthly timescale where temporal variations in the soil moisture are smoothed over the month. The performances of five streamflow forecasting models are compared:

- a '***base***' ANN model with candidate standard inputs (e.g. previous and forecast climate data, previous streamflow) and no soil moisture inputs;
- three models which use the three different soil moisture predictors separately:
  - '***base+GR4J SM***': ***base*** model + GR4J simulated soil moisture inputs;
  - '***base+ECV SM***': ***base*** model + remotely sensed ECV soil moisture inputs;
  - '***base+POAMA SM***': ***base*** model + POAMA forecast soil moisture inputs; and
- a model which includes all soil moisture data sets as inputs ('***base+all SM***').

As mentioned, for the purpose of this investigation, models are developed to forecast flow one month in advance at site A2390512 only. An artificial neural network (ANN) is used as the forecasting model and a Bayesian calibration method is employed such that the uncertainty in the streamflow forecasts can be quantified.

5.3 Data

As previously mentioned, in order to forecast future streamflows, information describing the initial catchment conditions and the effect of climate during the forecast period is required. In addition to soil moisture data, variables considered for representing initial catchment condition include previous streamflows, rainfall and evapotranspiration, as well as the Antecedent Precipitation Index (API) and groundwater levels. Rather than relying on climate indices to represent future climate, rainfall forecasts from the POAMA forecast system were used to represent the climate during the forecasting period, as it has recently been shown that, when used as predictors of future climate, forecasts from dynamical climate models may improve statistical streamflow forecasts when compared to the use of lagged climate indices only (Pokhrel et al., 2012).

For the purposes of the study carried out in this section, monthly forecast models were developed, requiring monthly input data and producing monthly forecasts. While daily data were available for many of the potential inputs, others were only available at greater time scales (e.g. groundwater) or daily data were only irregularly available over significant periods of the historical record (e.g. remotely sensed soil moisture). Furthermore, POAMA forecast data were difficult to handle at a daily time scale due to large file sizes and the great number of files which required downloading and processing (each daily file is approximately 300MB in comparison to ~20MB for monthly forecast files). Although monthly input data may not accurately account for within month variability or its influence on streamflow, a monthly model was considered appropriate since forecasts of monthly streamflow volumes were of primary interest.

Streamflow at A2390512 over the upcoming month is dependent on climate and catchment conditions both within and upstream of this catchment. When selecting potential input data for developing the statistical streamflow forecasting models, the following constraints were applied:

- The historical record must be sufficiently long at a monthly time step;
- Data should be relatively stationary; and
- All variables must be currently measured and available in near real-time (to be useful for operational purposes).

Data common to all modelling approaches applied in this report (i.e. flow, rainfall and PET) are outlined in Section 2. However, any preprocessing of this common data and any additional data considered as potential input information are outlined herein.

5.3.1 Flow

As described in Section 2.1, daily flow data are available from four different gauging stations within the Drain M catchment (see Figure 4). Flows at sites A2390512, A2390514 and A2390519 have been recorded from 1971 until the present, while flows have been recorded at A2390541, the outlet of Bool Lagoon, since 1985. Daily flows were converted to monthly totals, providing between 480-487 valid monthly records at A2390512, A2390514 and A2390519 and a total of 316 monthly records at A2390541. However, the information content of the complete flow time-series is relatively small, with regard to both the predictors and the predictand, as a result of the large number of zero flows recorded primarily in the summer and autumn months (see Figure 5). Therefore, to increase the influence of non-zero flows in the dataset and improve model calibration capabilities, it was considered sensible to forecast flows from the months June to November (winter and spring) only, when non-zero flows were dominant (and the results are of most interest). The removal of summer-spring flows resulted in 241-245 records at A2390512, A2390514 and A2390519 and 157 records at A2390541. Boxplots of total monthly flow for months June to November are shown in Figure 26 for each of the flow gauging stations. The monthly flow distributions for station A2390512 are shown in Figure 27 for all months and for June – November only. As can be seen in this figure, the distribution is minimally affected, if at all, by the removal of months December – May for flows above approximately 5-10 GL per month.

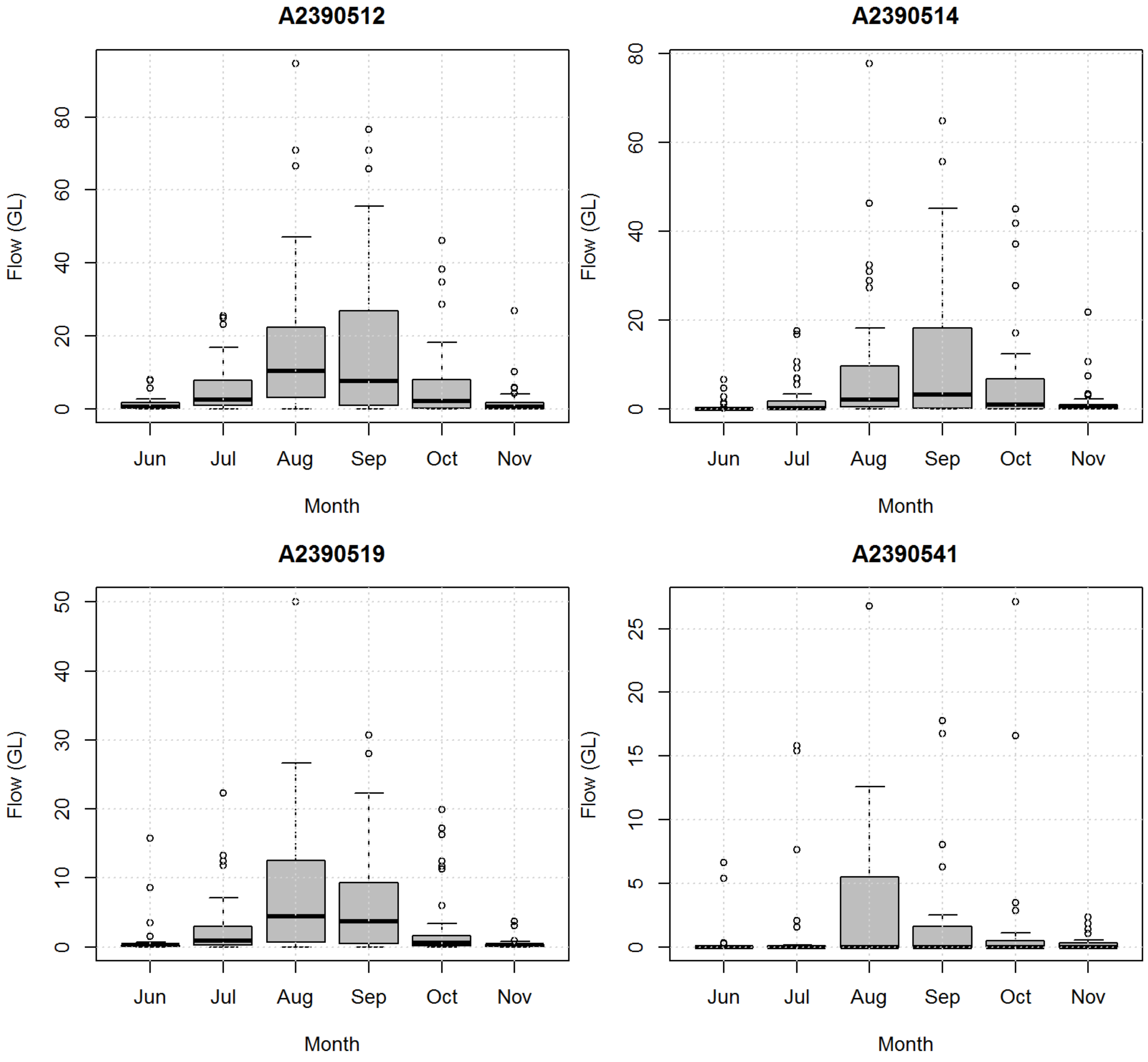


Figure 26 Total monthly flow boxplots by month (June – November)

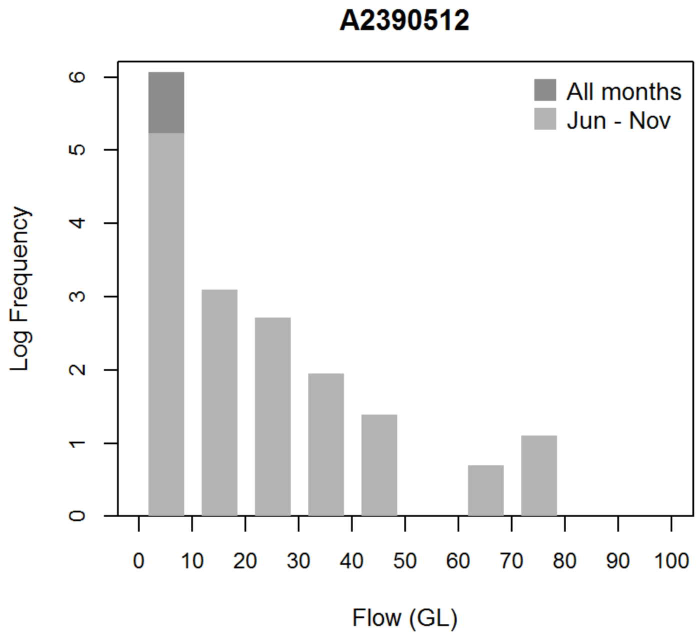


Figure 27 Overall monthly flow distributions

5.3.2 Rainfall and Evapotranspiration

As described in Section 2.2, daily rainfall and PET data from various stations around the Drain M catchments (see Figure 4), obtained from the SILO database, were combined using a Thiessen polygon approach to produce one time series each of rainfall and evapotranspiration for each catchment (A2390512, A2390514 and A2390519). Rainfall and PET data for each subcatchment are available from 1889 until the present. Similar to the flow data, daily time series were converted to monthly totals and the months November to April were removed (in order to forecast June - November flows with a lead time of one month), resulting in 738 data points each for rainfall and evapotranspiration for each subcatchment.

By only considering data between May and October, the overall monthly rainfall distributions become more normally distributed, as shown in Figure 28 (a) for subcatchment A2390512. The same plots for PET, shown in Figure 28 (b), also for subcatchment A2390512, illustrate that the distribution of May to October data is smoother than when all months are considered. Similar changes to the rainfall and PET distributions were observed for catchments A2390514 and A2390519 when November to April data were removed; however these are not shown here for the purpose of brevity.

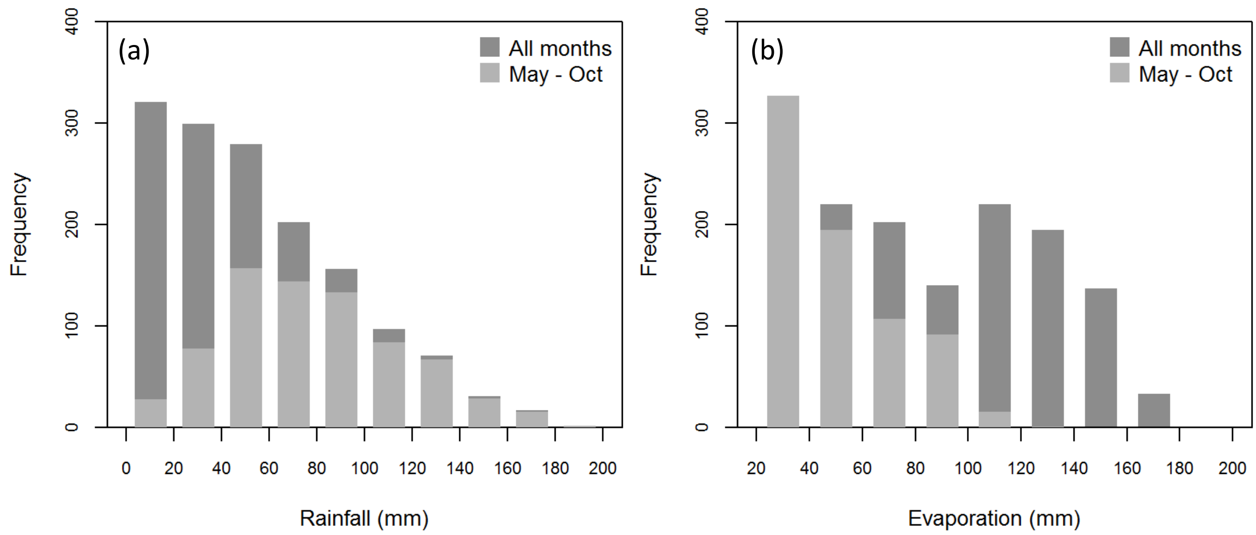


Figure 28 Monthly (a) rainfall and (b) PET distributions for all months and for May-October only. Distributions are for catchment A2390512.

5.3.3 Antecedent Precipitation Index

The API is a popular index used for representing initial catchment wetness in hydrological practice (Brocca et al., 2008) and is given by:

$$API_i(t) = \sum_{j=1}^{i} P_{t-j} k^{-j} \tag{12}$$

where *i* is the number of antecedent days included in the calculation, *k* is a recession coefficient and $P_{t\text{-}j}$ is the precipitation for day *t-j*. The recession coefficient *k* represents the "memory" of a catchment by decaying the effect of accumulated rainfall at each time step. The theory is that earlier precipitation should have less influence on present streamflow response than recent precipitation. While *k* is typically set between 0.8 and 0.98 (Heggen, 2001), a value of *k*=0.99 was used for each catchment based on a recession analysis of the daily flow data (see Figure 29), under the assumption that the effect of antecedent precipitation on streamflow would decay at the same rate as the recession limb of a hydrograph during periods without rainfall.

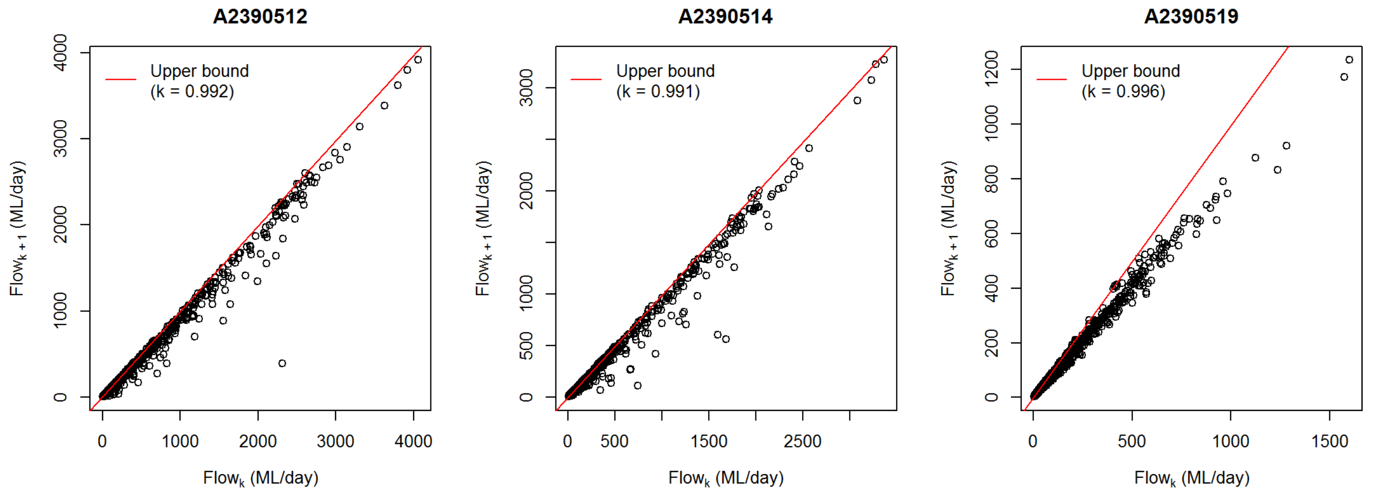


Figure 29 Recession analysis of daily flows for each catchment

As rainfall data are available from 1889 until the present, daily API values for each catchment were also computed for this period. The daily values were then converted to monthly values by taking the average over each month and again, November to April were discarded, giving a total of 738 values for each catchment.

5.3.4 Groundwater

A number of groundwater monitoring wells are located in and around the Drain M catchments, as can been seen in Figure 30. However, to be considered as potentially useful from a statistical model building point of view, records were required to be sufficiently long at a monthly time scale. As such, only wells with records beginning prior to the year 2000 were considered (providing a minimum of 72 monthly data points). Additionally, for operational purposes, data need to be available in near real-time; therefore, only groundwater observations from wells where telemetry is used were considered. This resulted in only three groundwater wells from which potentially useful information could be obtained. These were:

- SMT020 – with observations available approximately 6 monthly from 1980-1993 and approximately 3 monthly from 1993-present.
- CMM079 – with observations available approximately 6 monthly from 1977-1979 and approximately 3 monthly from 1994-present.
- MTM060 – with observations available approximately 3 monthly from 1978-1982 and from 1996-present.

Groundwater data were extracted from the South Australian groundwater database (https://www.waterconnect.sa.gov.au/Systems/GD/Pages/default.aspx) and the recorded observations were linearly interpolated (during periods when there were at least 6 monthly observations available) to provide monthly average groundwater depths.

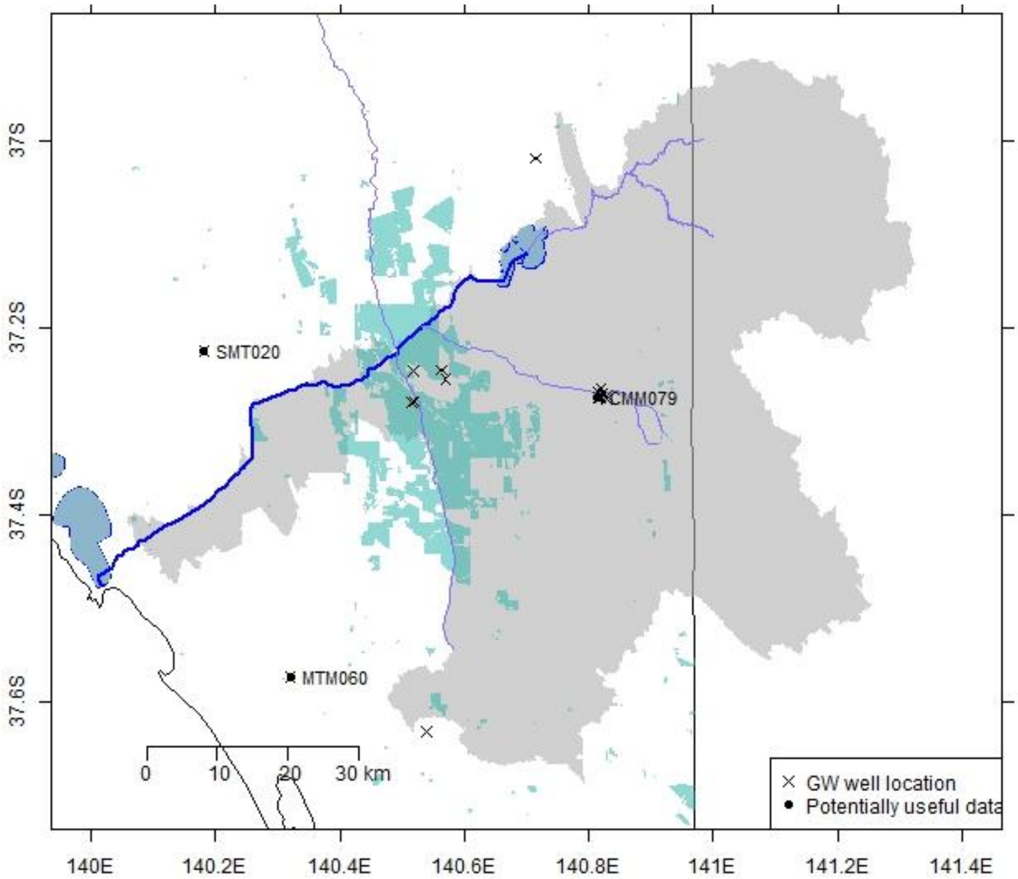


Figure 30 Groundwater wells located in and around the Drain M catchment

As there are significant periods of missing data for wells CMM079 and MTM060, the Hydrograph Analysis - Rainfall and Time Trend (HARTT) software (Ferdowsian et al., 2001) was used in the attempt to fill in these periods. This method involves the use of a regression model to estimate depth of groundwater below the ground surface:

$$\text{Depth}_t = k_0 + k_1 \, \text{AMRR}_{t-L} + k_2 \, t \qquad (13)$$

where AMRR is the accumulative monthly residual rainfall (accumulated deviations from average rainfall), t is the number of months since observations commenced, L is length of time lag (in months) between rainfall and its impact on groundwater, and k_0, k_1 and k_2 are parameters to be estimated. Parameter k_0 is approximately equal to the initial depth to groundwater, k_1 represents the impact of above- or below-average rainfall on the groundwater level, and k_2 is the trend rate of groundwater rise or fall over time. The value of L is estimated by selecting the value that results in the highest R^2 for the regression (Ferdowsian et al., 2001).

As can be seen in Figure 31, the HARTT model provided a good estimate of the measured groundwater depth for well CMM079 from 1994 onward and, therefore, it was considered that this method was appropriate for filling in the groundwater record from this well between 1979 and 1994. However, this was not the case for other wells in the region, and as such, could not be used to extend the length of any other records for wells where telemetry is used. This can be seen in Figure 32 which shows the monthly groundwater levels for well MTM060 together with the levels estimated by the HARTT model. Consequently, only groundwater data from wells CMM079 and SMT020, which had limited missing data as shown in Figure 33, were considered as potential inputs to the ANN models.

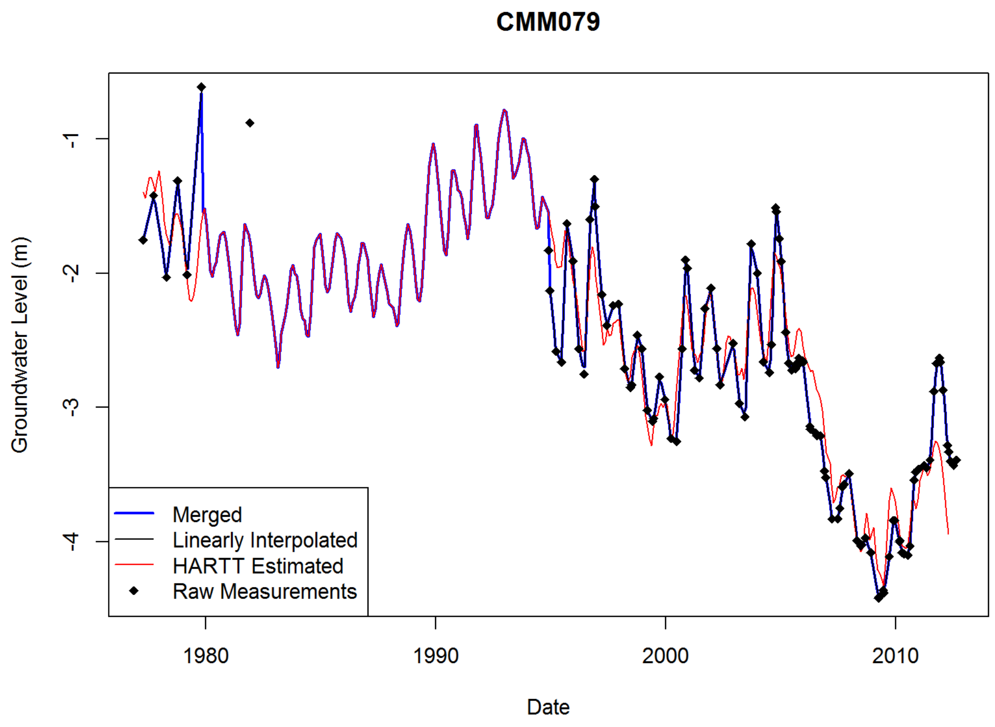


Figure 31 CMM079 average monthly groundwater levels

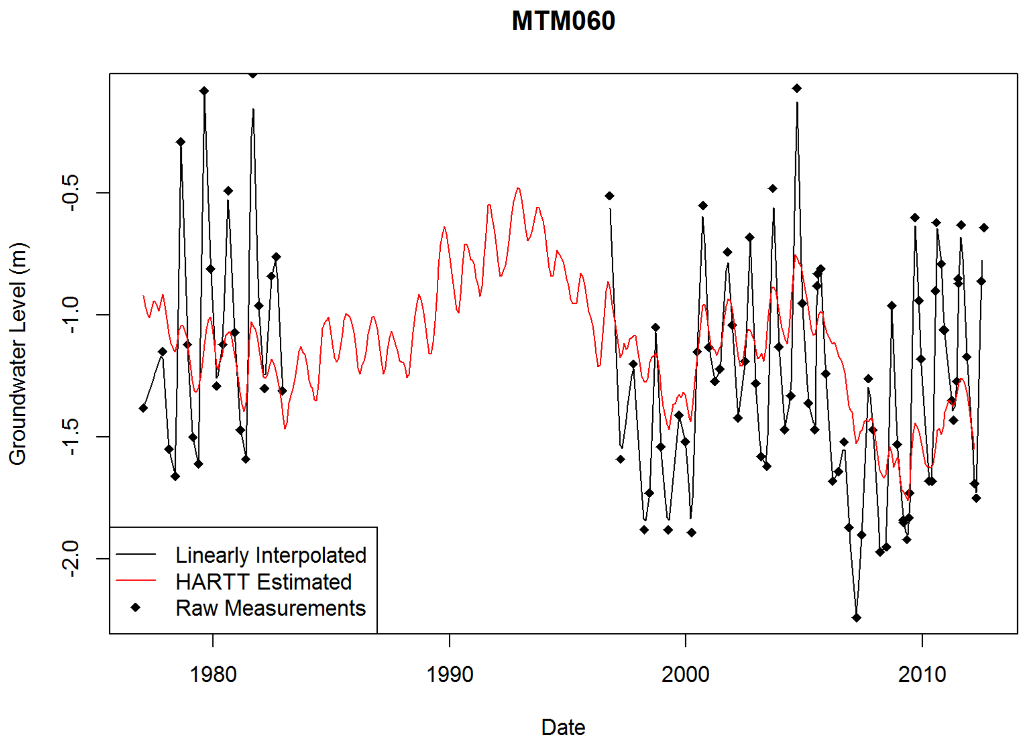


Figure 32 MTM060 average monthly groundwater levels

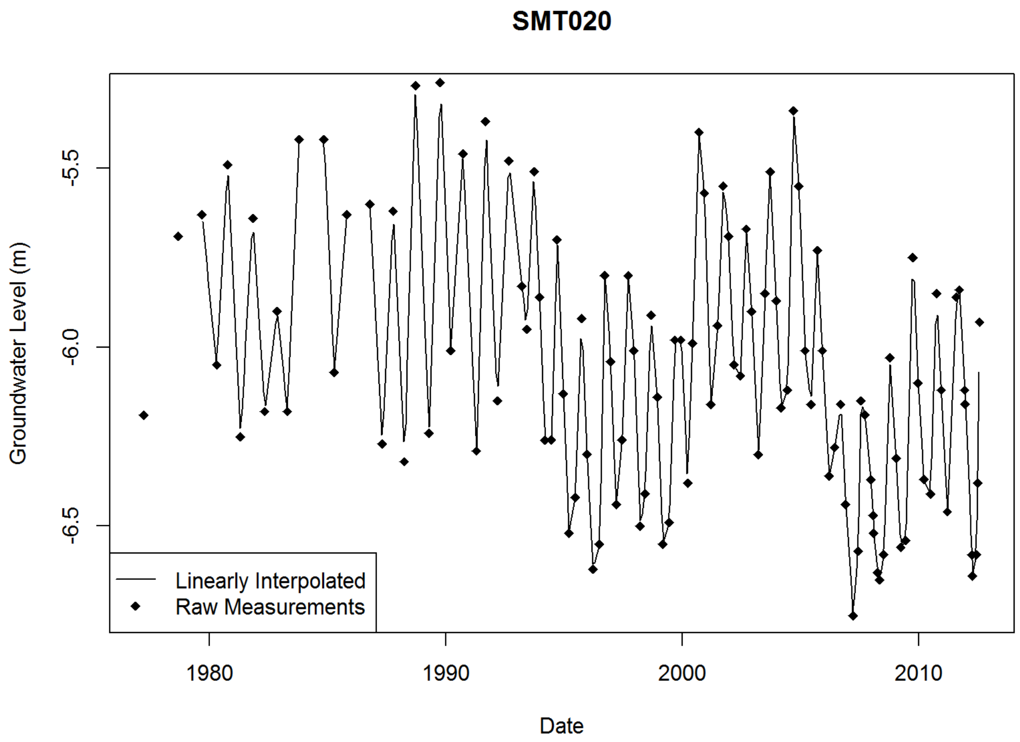


Figure 33 SMT020 average monthly groundwater levels

5.3.5 POAMA Rainfall Forecasts

Rainfall forecasts from the Australian Bureau of Meteorology's (BoM) seasonal forecast system, POAMA, were used to represent the effect of climate during the forecast period. POAMA is a dynamical climate forecasting system designed to produce multi-week to seasonal forecasts of climate for Australia based on a coupled ocean/atmosphere model and ocean/atmosphere/land observation assimilation systems.

POAMA-2 consists of both a seasonal and a monthly/multi-week forecast system using version 2.4 of the model. Both of these systems are multi-model ensembles consisting of three different model configurations. Furthermore, each configuration is used to generate an ensemble of 10 forecasts by perturbing initial conditions, resulting in a 30-member ensemble. POAMA-2 hindcast and realtime data are available as an experimental product from the BoM[1]. For developing the ANN models, monthly hindcasts from the monthly/multi-week forecast system were used. These are available from January 1981 until December 2011 and are produced on the 1st, 11th and 21st of the each month during this period. Hindcasts made on the 21st of each month were chosen for model calibration, as it was considered that these would produce the most accurate forecasts for the next month (i.e. a mean rainfall forecast for June made on the 21st of May would likely be more accurate than a forecast made on the 1st of May, for example).

POAMA produces large spatial scale predictions over Australia (see Figure 25), which do not capture the large degree of spatial variability in rainfall. Therefore downscaling of the large scale POAMA rainfall hindcasts to rainfall at a more local scale was carried out. The statistical downscaling method detailed in Shao and Li (2013) was used for this, resulting in rainfall hindcasts at each of the rain gauges shown in Figure 34. The downscaled 30-member ensembles were reduced to three time series at each rain gauge: the minimum, mean and maximum of the

[1] While the real-time forecasts produced by POAMA-2 are based on the observations available at the time, the hindcasts are based on a clearer picture of the weather conditions which is derived after all data and observations have arrived (a re-analysis). As such, the hindcasts are based on slightly different initial conditions to those that are used to create a forecast and, consequently, the error of the forecasts will be higher due to the greater error associated with the initial conditions.

ensembles. As can be seen in Figure 35, the mean rainfall hindcasts at station 26000 provide a reasonable estimate of the observed rainfalls at this station, while the upper and lower hindcast bounds capture the majority of observed rainfalls. The same result was seen for the other three stations; however, results are not shown here for the purpose of brevity.

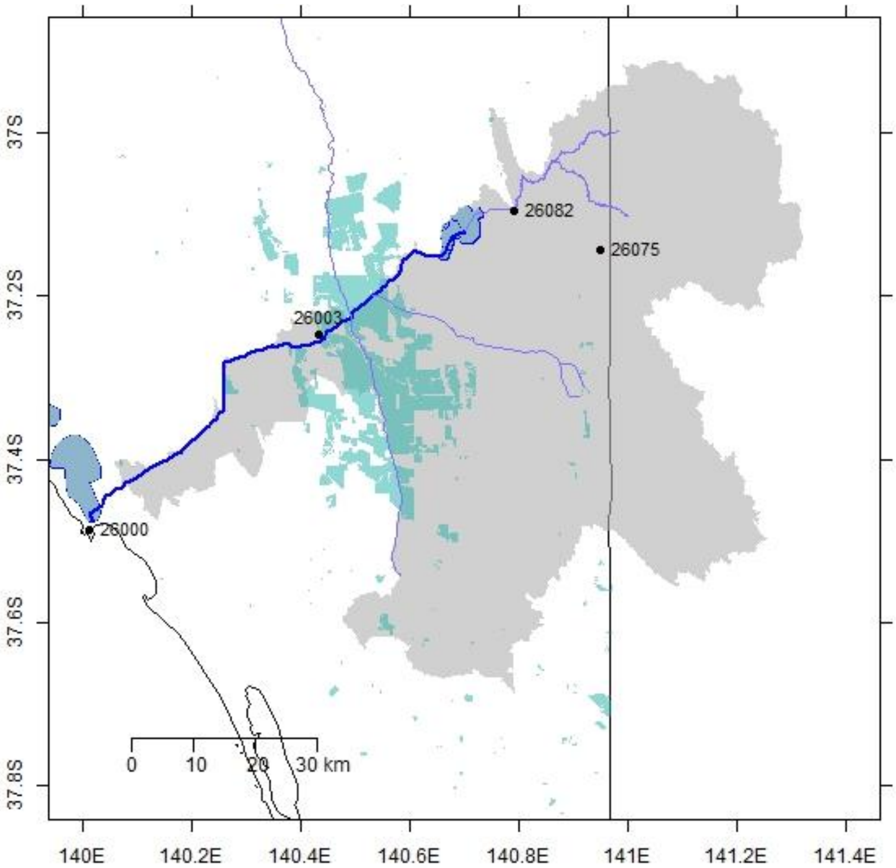


Figure 34 Rainfall station locations

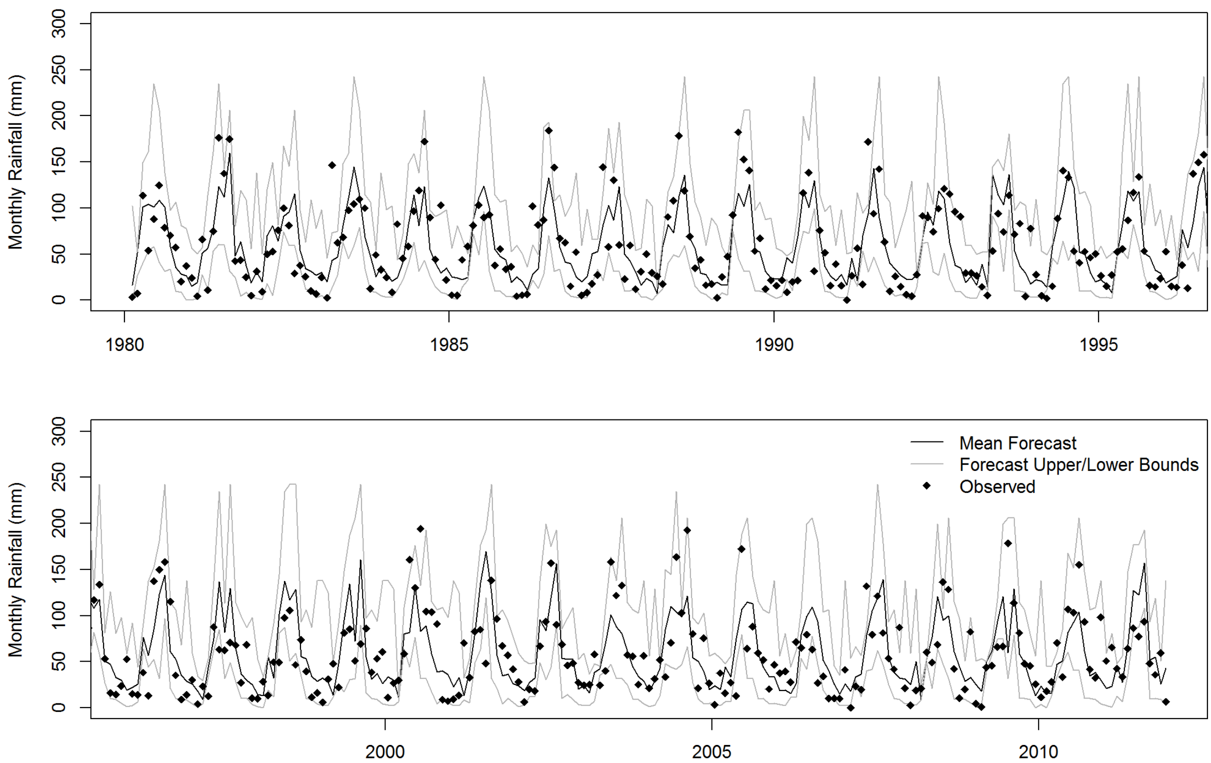


Figure 35 POAMA rainfall forecasts at station 26000 in comparison to observed rainfalls

5.3.6 Soil Moisture

As previously discussed, three different sources of soil moisture data were considered as direct inputs to the ANN models developed. Details of each of these data sets are given in the following subsections.

5.3.6.1 Remotely sensed soil moisture

There are a number of remotely sensed global soil moisture datasets available from various satellites (both passive and active), the longest of these dating back to 1978. While these records vary in both length and quality, they all provide soil moisture estimates for the top few centimetres, typically at a rather coarse spatial resolution of 25 km–50 km.

Soil moisture has been recognised as an ECV for which satellite data are required to meet the needs of the climate change community. In 2010, the European Space Agency (ESA) Programme on Global Monitoring of Essential Climate Variables established the Soil Moisture Climate Change Initiative (CCI) project in order to meet this need. The main objective of this project was to create a long-term consistent satellite soil moisture time series, based on active and passive data (De Jeu et al., 2012). To do this, two extensively validated soil moisture datasets were selected to create a harmonized dataset:

- An active data set generated by the Vienna University of Vienna (TU Wien) based on observations from the C-band scatterometers on board of ERS-1, ERS-2 and METOP-A.
- A passive data set generated by the VU University Amsterdam in collaboration with NASA based on passive microwave observations from Nimbus 7 SMMR, DMSP SSM/I, TRMM TMI and Aqua AMSR-E.

Version 1 of the merged dataset, which covers the 32 year period from 1978 to 2010, was used in this study. While this dataset does not meet the constraint that it must be currently available in near real-time, a soil moisture product derived from the Advanced SCATterometer (ASCAT) on-board the METOP-A satellite is distributed in near real-time by the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) (http://www.eumetsat.int/website/home/Data/Products/Land/index.html). This is one of the primary datasets which makes up the merged set. Furthermore, regular yearly updates of the merged dataset are planned, with version 2 now available from http://www.esa-soilmoisture-cci.org, covering the 35 year period from 1978 to 2013. In future this dataset will potentially include soil moisture measurements from the recently launched Soil Moisture Active Passive (SMAP) mission, which will provide soil moisture data at a higher spatial resolution (although these data will not be useful for operational purposes, they may be useful for further testing and validation of the models developed or recalibration if required).

The merged ECV soil moisture dataset, henceforth referred to as ECV SM, is available at a spatial resolution of approximately 25 km (0.25°), a temporal resolution of approximately one day (although temporal resolution is irregular over the period of available data, particularly prior to 1998 when resolution can be in the range of one day to two weeks), and representing a soil layer depth of 0.5 – 2 cm. The spatial grid of ECV SM data over the Drain M catchment is shown in Figure 36. As there are a number of ECV SM grid points, and hence observations, either inside or adjacent to the catchment boundaries (particularly in the case of catchment A2390514), it was considered undesirable to simply convert these to estimates at the catchment centroids, as this would lose any spatial variation in soil moisture captured in the observations. Therefore, the overall catchments were divided into the subcatchments shown in Figure 37. The details of these subcatchments, which were identified using a LiDAR derived 10 m DEM (Wood and Way, 2011), are given in Table 6. The ECV SM data were converted to monthly average values and spatially interpolated to produce monthly soil moisture estimates at the centroid of each of the subcatchments shown in Figure 37. The summer and autumn months were then removed to be consistent with other input data, resulting in 191 monthly data points between 1978 and 2010.

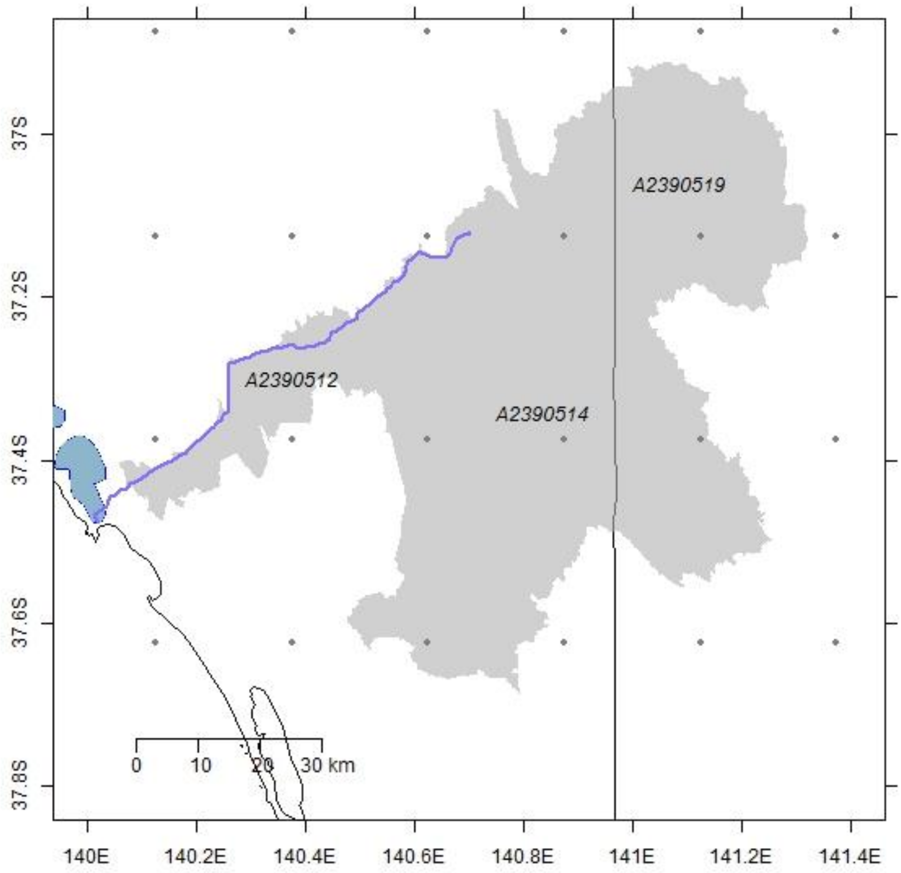


Figure 36 Spatial grid of ECV soil moisture over the Drain M catchments

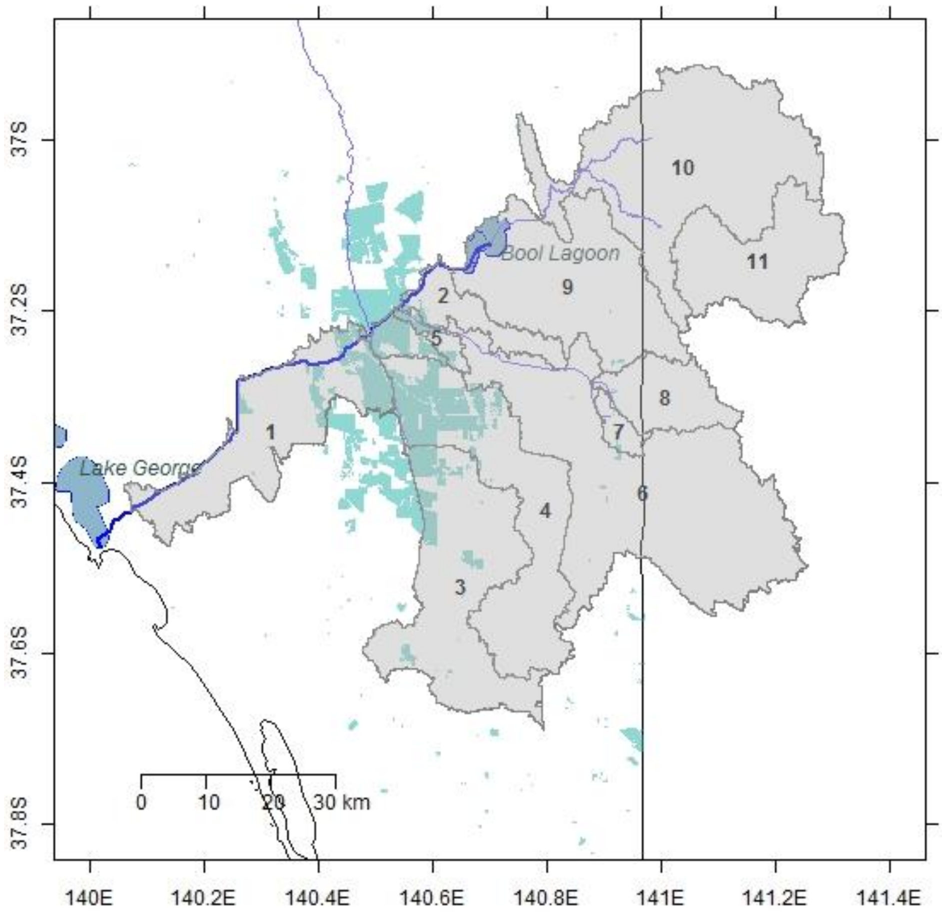


Figure 37 Subcatchments used for soil moisture estimates

Table 6 Subcatchment details

ID	Name	Area (km^2)	Mean Annual Rainfall (mm)
1	A2390512	383	676
2	A2390514	154	634
3	A2390515	493	735
4	A2391001	483	686
5	DrainC	29	642
6	A2390516	858	658
7	A2390537	33	638
8	A2390536	149	643
9	A2390541	489	618
10	A2390519	701	598
11	A2391076	302	626

5.3.6.2 POAMA forecast soil moisture

The POAMA-2 forecast system described in Section 5.3.5 also produces forecasts (and hindcasts) of soil moisture at the same temporal and spatial resolutions as the forecast rainfall. The POAMA-2 system uses a simple, single-layer bucket model, with a field capacity of 15 cm, to represent soil moisture (Hudson et al., 2013; Manabe and Holloway, 1975). The soil moisture forecasts, henceforth referred to as POAMA SM, were pre-processed in the same manner as the POAMA rainfall forecasts. Due to the coarse spatial resolution of the data (shown in Figure 25), and limited ability of the POAMA system to represent the soil moisture process, the downscaling method of Shao and Li (2013) was also used to downscale the soil moisture forecasts to a more local scale. While this method was developed for downscaling precipitation, rather than soil moisture, it was considered appropriate under the assumption that the same atmospheric conditions that drive precipitation would also influence soil moisture, given that precipitation is the main driver of variations in soil moisture (although this assumption does not account for the influence of previous precipitation on soil moisture). The coarse resolution forecasts were downscaled to produce soil moisture forecasts at the centroids of each of the sub-catchments shown in Figure 37 using the ECV SM estimates as the local data. While the remotely sensed data may still not be considered local scale data (compared to, for example, in-situ soil moisture probes), these were the only observations available at a finer spatial resolution than the POAMA data. After this downscaling process, the POAMA SM data provide forecasts of soil moisture in the top few (~2) centimetres of the soil profile only, similar to the ECV SM data. Removing summer and autumn months from these data resulted in 186 monthly soil moisture forecasts at each subcatchment centroid between 1980 and 2010.

5.3.6.3 Soil moisture index from GR4J

Similar to Anctil et al. (2004), a soil moisture index (SMI) was simulated using the standard four parameter GR4J CRR model described in Section 3. The SMI is the daily level in the 'production store', S (see Figure 10). This represents the soil water content inside the conceptual soil reservoir and is thus analogous to total soil moisture (unlike the ECV and POAMA SM data, which represent surface soil moisture only). The SMI is balanced according to the relative magnitudes of rainfall and PET and can be depleted by percolation. The GR4J models used to simulate the SMI series for each subcatchment were calibrated to daily flows at A2390512 between 1971 and 2012, with a rolling 10 year period used for calibration (see Section 4). The final year in each calibration period was used to provide the SMI time series, as this is expected to produce the best possible estimate of the internal model parameters, including the daily level of the production store (i.e the SMI). The daily SMI was extracted and converted to monthly average values to provide a monthly SMI time series. The removal of summer and autumn months resulted in 228 monthly data points between 1974 and 2011.

Shown in Figure 38 are time series plots of the monthly SMI, henceforth referred to as GR4J SM, in comparison to the ECV SM and POAMA SM for subcatchment A2390512 (this is the only catchment where these three soil moisture data sets were comparable, as A2390512 was not subdivided for the purpose of estimating ECV SM and POAMA SM). Although the ECV SM and POAMA SM data represent surface soil moisture only, while GR4J SM data represent total soil water storage (i.e. the three datasets are not consistent in their representations of soil moisture), it can be seen that there is reasonable agreement between the temporal evolution of soil moisture, as represented by the three different datasets. Furthermore, while there are some discrepancies between the ECV and POAMA SM datasets, their fairly close agreement in both timing and magnitude suggests the POAMA SM data should provide reasonable forecasts of surface soil moisture over the upcoming month. As can be seen by the coefficient of variation (CV) values also given in Figure 38, these two datasets display a similar degree of variation (as would be expected, since they are both derived from the ECV SM data), while there is greater variation in the GR4J SM data. This is possibly due to the more local scale of the GR4J SM data (the ECV and POAMA SM data have been smoothed in space) or their representation of the total soil water store, rather than the upper few centimetres only.

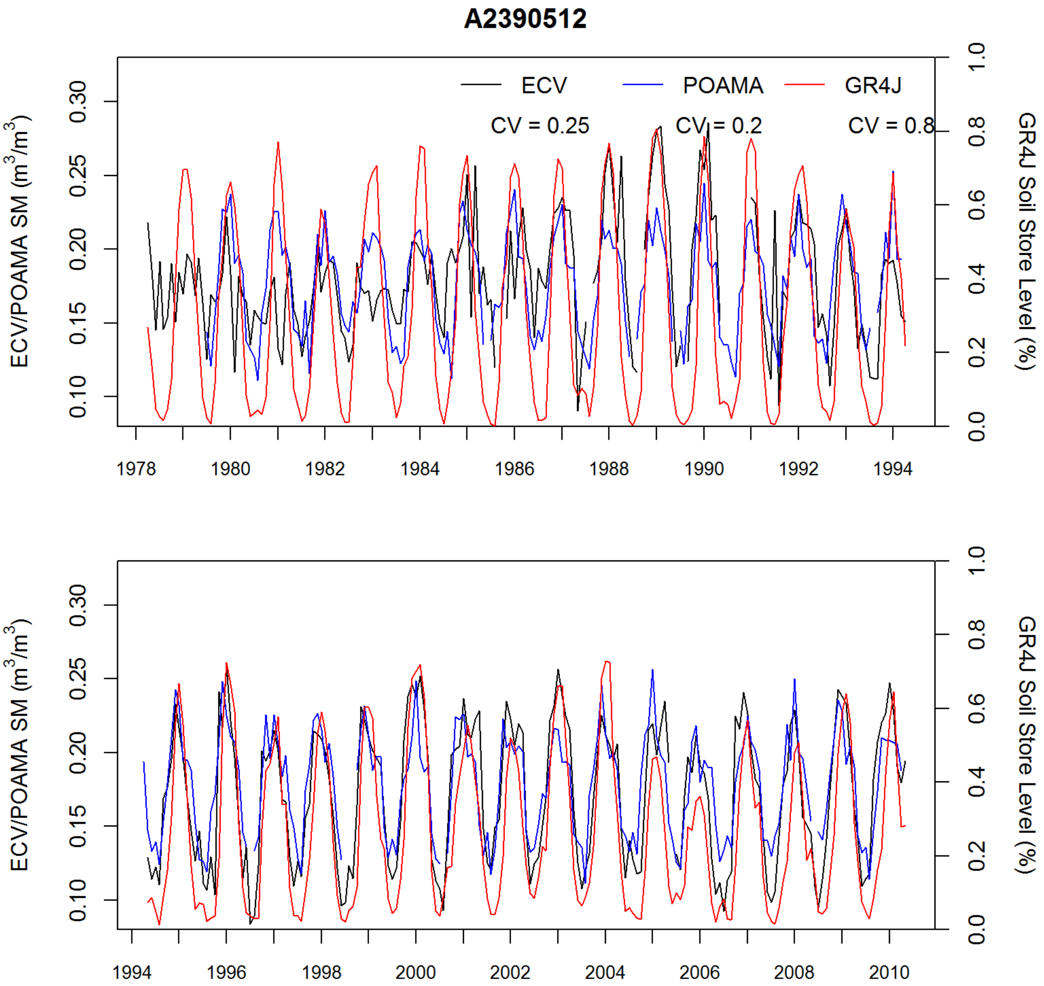


Figure 38 Comparison of 3 soil moisture data sets: ECV SM, POAMA SM and GR4J SM

5.4 Modelling methodology

Artificial neural networks (ANNs) are an extremely versatile type of data-driven model which have become widely adopted for streamflow forecasting over the past two decades (see Maier et al. (2010)). Their flexible model structure enables them to capture complex and nonlinear input-output relationships from data without any restrictive assumptions about the functional form of the underlying process, giving them more general appeal than less flexible data-driven models. Feedforward multi-layer perceptrons (MLPs), the type of ANN most commonly used for hydrological modelling (Maier et al., 2010), were used in this study. A typical MLP structure, with a single hidden layer and one output, is shown in Figure 39. The input layer receives information from each of the *K* input variables, which is then passed through the network in a feedforward direction to subsequent layers in the network. This is done via weighted connections, where each node in one layer is connected to every node in the next layer. As such, the input information is redistributed across all of the hidden layer nodes where it is then processed. The complexity of a MLP can easily be increased (decreased) through the addition (removal) of hidden layer nodes. At each of the hidden and output layer nodes, the weighted information from the previous layer is summed together with a weighted offset, or bias, and then transformed by the node's transfer, or activation, function. The transfer function may be any continuous differentiable function, but is commonly sigmoidal at the hidden layer nodes and linear at the output layer. The connection and bias weights are the free parameters of the model and, during calibration or training of a MLP, optimal values for these weights are sought such that the outputs produced by the ANN approximate the observed training data well.

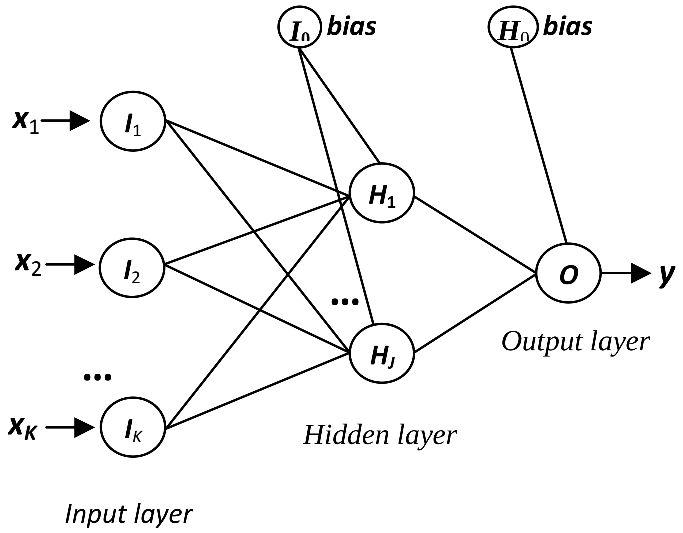


Figure 39 Typical structure of a MLP

5.4.1 Input Variable Selection

The appropriate selection of inputs is one of the most critical steps in the development of ANN models. The model inputs, together with the response variable, contain all of the information necessary for characterising the underlying process, given that no descriptions of the physical processes are explicitly included in the model. For streamflow forecasting, the database of potential or candidate inputs, which is defined based on a priori information about the physical processes and by whichever data are available, usually includes observations of the predictors (e.g. those which represent initial catchment conditions and the effect of climate) at different locations and time lags, as well as lagged observations of the streamflow being modelled. Consequently, the number of potentially important inputs can be quite large; however, given their correlated nature, many may be redundant. Furthermore, some potential inputs (e.g. coarse-scale spatial data or point measurements in particular locations) may provide little or no information about the rainfall-runoff process, making them irrelevant to the problem. The inclusion of irrelevant and/or redundant inputs only adds noise and complexity into the model, while the omission of relevant input variables, on the other hand, results in part of the output behaviour remaining unexplained.

Due to the non-linearity and inherent complexity associated with hydrological systems, well established analytical input variable selection (IVS) methods, such as correlation analysis, are typically inadequate for hydrological modelling applications (May et al., 2011). As such, an automatic IVS algorithm was used in this study to select the appropriate inputs for forecasting flows in Drain M. Such methods systematically search the potential input space (i.e. the candidate input pool) for individual inputs or combinations of inputs that are relevant (and non-redundant) to the model based on some defined evaluation criteria (Blum and Langley, 1997; Dash and Liu, 1997; Guyon and Elisseeff, 2003; Liu et al., 2010). It has been shown that these algorithms are able to improve prediction accuracy and produce more parsimonious models in numerous studies when compared to more conventional IVS methods (e.g. correlation analysis) or the inclusion of all available input data (see Galelli et al. (2014) for a discussion on the use of automatic IVS algorithms in environmental modelling).

5.4.1.1 IVS algorithm

The GA-ANN IVS approach described in Galelli et al. (2014) was used in this study to select the "optimum" set of inputs for each of the models developed. This method combines a genetic algorithm (GA) search procedure (Goldberg, 1989) with a simple, 1-hidden node ANN model, where the aim is to identify the subset of inputs which yields the best ANN model performance. The GA-ANN algorithm was selected for IVS in this study as preliminary investigations showed that an ANN with a single hidden layer and one hidden node was capable of accurately simulating flows at A2390512 and, therefore, it can be considered tailored to this study in that the selected inputs are those that optimise the performance of the chosen induction algorithm (and should therefore yield high forecast accuracy).

The GA-ANN algorithm can be summarised by the following steps:

1. Define the search space by identifying the set of potentially important inputs.
2. Initialise the population of solutions by sampling binary chromosomes at random from within the defined search space.
3. Evaluate the fitness of each chromosome by training the ANN model for the combination of inputs represented and evaluating the objective function.
4. Select the fittest chromosomes using `tournament selection', whereby pairs of chromosomes are randomly competed against one another and the chromosome with the best fitness is selected.
5. Apply crossover and mutation operators to the selected chromosomes to form the next generation of solutions. Crossover involves the exchange of information, beyond a randomly selected crossover point, between randomly paired chromosomes according to some crossover probability. Mutation of a chromosome simply involves the occasional random alteration of the value of a gene (i.e. a 0 will become a 1 and vice versa) according to a small mutation probability.
6. Repeat steps 3-5 until there is no improvement in the best overall fitness.
7. Terminate the algorithm.

The objective function used to determine whether one subset of inputs is better (fitter) than another was the out-of-sample AIC. This was computed using k-fold cross-validation, where the number of folds, k, was equal to 5. This objective function was also used as a stopping criterion to terminate the GA-ANN algorithm (see step 6 above).

The stochastic GA search strategy employed by the GA-ANN IVS method is more adept at finding (near) globally optimum input subsets than methods which employ local search strategies to successively add or remove candidate input variables according to their perceived relevance (see Galelli et al. (2014) for a more thorough discussion of the advantages and disadvantages of different IVS methods). However, it is still not possible to guarantee that a globally optimal set of inputs has been found by the algorithm. As such, for each of the models developed, the GA-ANN method was run 10 times to ensure that locally optimal solutions were not being converged upon and that the best set of inputs was being selected. This is a time consuming process; however, since it need only be carried out once during the model development process, it was considered worthwhile for achieving the greatest forecast accuracy.

5.4.1.2 Candidate input pool

To generate the candidate input pool for the ***base*** model (with no inclusion of soil moisture data), each of the more traditional predictors for representing the initial catchment condition (i.e. previous rainfall, evaporation, streamflow, etc.) were lagged by up to two months, while 1-month ahead POAMA forecast rainfalls were represented by a minimum, mean and maximum value, as described in Section 5.3.5. For the three models that included a single source of soil moisture data (***base+GR4J SM***, ***base+ECV SM*** and ***base+POAMA SM***), the respective candidate input pools were restricted to only include the more traditional catchment wetness predictors and POAMA rainfall forecasts that were found to be optimal for the ***base*** model, with the addition of lagged values of the soil moisture data of interest (or minimum, mean and maximum values in the case of the POAMA forecast soil moisture data).

The candidate input pool for the ***base+all SM*** model was similarly restricted such that the GA-ANN algorithm could only select from the inputs already found to be optimal for the previous four models. This was done so that any direct gains associated with the inclusion of particular soil moisture inputs, or combination thereof, in addition to or in place of the more traditional rainfall-runoff inputs could be identified without the confusion of having completely different combinations of inputs in each model which could result due to the high degree of correlation between the candidate inputs and the fact that the GA-ANN algorithm can not guarantee a globally optimal solution. Furthermore, the smaller the candidate input pool, the easier it is to find optimal combinations of the soil moisture inputs to complement or replace the inputs used in the ***base*** model. Inputs included in the respective candidate input pools for each model are given in Table 7.

Table 7 Candidate Inputs for each statistical streamflow model

base model***:*** a total of 57 candidate inputs: • Flow_A2390512$_{t-i}$, Flow_A2390514$_{t-i}$, Flow_A2390541$_{t-i}$, Flow_A2390519$_{t-i}$, • Rain_A2390512$_{t-i}$, Rain_A2390514$_{t-i}$, Rain_A2390519$_{t-i}$, • Evap_A2390512$_{t-i}$, Evap_A2390514$_{t-i}$, Evap_A2390519$_{t-i}$, • API_A2390512$_{t-i}$, API_A2390514$_{t-i}$, API_A2390519$_{t-i}$, • GW_SMT020$_{t-i}$, GW_CMM079$_{t-i}$, • POAMA_Rain_26000_min$_{t+1}$, POAMA_Rain_26000_mean$_{t+1}$, POAMA_Rain_26000_max$_{t+1}$, • POAMA_Rain_26003_min$_{t+1}$, POAMA_Rain_26003_mean$_{t+1}$, POAMA_Rain_26003_max$_{t+1}$; • POAMA_Rain_26075_min$_{t+1}$, POAMA_Rain_26075_mean$_{t+1}$, POAMA_Rain_26075_max$_{t+1}$; • POAMA_Rain_26082_min$_{t+1}$, POAMA_Rain_26082_mean$_{t+1}$, POAMA_Rain_26082_max$_{t+1}$; where i = 0, 1, 2
base + ECV SM: a total of 40 candidate inputs: • 7 ***base*** model inputs (see Table 9), • ECV_SM_A2390512$_{t-i}$, • ECV_SM_A2390514$_{t-i}$, • ECV_SM_A2390515$_{t-i}$, • ECV_SM_A2391001$_{t-i}$, • ECV_SM_DrainC$_{t-i}$, • ECV_SM_A2390516$_{t-i}$, • ECV_SM_A2390537$_{t-i}$, • ECV_SM_A2390536$_{t-i}$, • ECV_SM_A2390541$_{t-i}$, • ECV_SM_A2390519$_{t-i}$, • ECV_SM_A2391076$_{t-i}$; where i = 0, 1, 2
base + POAMA SM: a total of 40 candidate inputs: • 7 ***base*** model inputs, • POAMA_SM_A2390512_min$_{t+1}$, POAMA_SM_A2390512_mean$_{t+1}$, POAMA_SM_A2390512_max$_{t+1}$, • POAMA_SM_A2390514_min$_{t+1}$, POAMA_SM_A2390514_mean$_{t+1}$, POAMA_SM_A2390514_max$_{t+1}$, • POAMA_SM_A2390515_min$_{t+1}$, POAMA_SM_A2390515_mean$_{t+1}$, POAMA_SM_A2390515_max$_{t+1}$, • POAMA_SM_A2391001_min$_{t+1}$, POAMA_SM_A2391001_mean$_{t+1}$, POAMA_SM_A2391001_max$_{t+1}$, • POAMA_SM_DrainC_min$_{t+1}$, POAMA_SM_DrainC_mean$_{t+1}$, POAMA_SM_DrainC_max$_{t+1}$, • POAMA_SM_A2390516_min$_{t+1}$, POAMA_SM_A2390516_mean$_{t+1}$, POAMA_SM_A2390516_max$_{t+1}$, • POAMA_SM_A2390537_min$_{t+1}$, POAMA_SM_A2390537_mean$_{t+1}$, POAMA_SM_A2390537_max$_{t+1}$, • POAMA_SM_A2390536_min$_{t+1}$, POAMA_SM_A2390536_mean$_{t+1}$, POAMA_SM_A2390536_max$_{t+1}$, • POAMA_SM_A2390541_min$_{t+1}$, POAMA_SM_A2390541_mean$_{t+1}$, POAMA_SM_A2390541_max$_{t+1}$, • POAMA_SM_A2390519_min$_{t+1}$, POAMA_SM_A2390519_mean$_{t+1}$, POAMA_SM_A2390519_max$_{t+1}$, • POAMA_SM_A2391076_min$_{t+1}$, POAMA_SM_A2391076_mean$_{t+1}$, POAMA_SM_A2391076_max$_{t+1}$
base + GR4J SM: a total of 16 candidate inputs: • 7 ***base*** model inputs, • GR4J_SM_A2390512$_{t-i}$; • GR4J_SM_A2390514$_{t-i}$; • GR4J_SM_A2390519$_{t-i}$; where i = 0, 1, 2
base + all SM: – A total of 14 candidate inputs: • 7 ***base*** model inputs; • 2 additional ***base + ECV SM*** model inputs (see Table 9); • 2 additional ***base + POAMA SM*** model inputs (see Table 9); • 3 additional ***base + GR4J SM*** model inputs (see Table 9).

5.4.2 ANN development and characterisation of forecast uncertainty

Standard ANN development practices (e.g. see, for example, Maier and Dandy (2000)) were initially employed to develop deterministic ANN models for each of the input scenarios to provide initial weights that would subsequently be used for the Bayesian training approach employed (see details below). Computations were implemented using the R "nnet" package (http://cran.r-project.org/web/packages/nnet/), which enables the development of feedforward MLPs with a single hidden layer. A logistic transfer function was used at the hidden layer nodes while a linear transfer function was used at the output layer, as it has been shown that a one hidden layer network with this configuration of activation functions has universal approximation capabilities (Bishop, 1995). The "Nelder-Mead" (or downhill simplex) method, as implemented within the nnet package, was used for training the models. To avoid this algorithm becoming trapped in a poor local optimum, training was initialised 20 times using different random weights and the best resulting model was selected based on out-of-sample performance on a test dataset. Early stopping using the same test dataset for cross-validation was also employed to prevent overfitting of the training data. In addition, cross-validation using the test dataset was used to select the optimum number of hidden nodes for each model. ANN models with 1, 2, …, 6 hidden nodes were developed and the model which provided the best out-of-sample performance on the test dataset was selected. Each model was subsequently validated using an independent validation dataset not used in the ANN development process.

The above procedure required that the model development data be divided into training, testing and validation subsets. For each model, five years of validation data were selected such that a mixture of both relatively high and low flows were experienced during these years. Data from 2000 to 2004 were used for independent validation, while the remaining data were randomly divided into training and test sets with proportions of 80% and 20%, respectively. To do this, the data were randomly reordered multiple times until the statistics of the training and testing portions of the data were most similar. After accounting for the appropriate lags of the input and output variables, the number of monthly data available for training, testing and validation are given in Table 8.

Table 8 Number of data samples used for model development

Training	Testing	Validation
110	27	30

The MCMC Bayesian training approach developed by Kingston et al. (2005) was subsequently used to infer the posterior probability distribution of the ANN weights, $\mathbf{w} = \{w_1, \dots, w_d\}$, given the observed data, $\mathbf{y} = \{y_1, \dots, y_N\}$, thereby accounting for the uncertainty associated with these model parameters and the resulting uncertainty in the model forecasts. An additional advantage of this method is that it does not require a test data set for cross validation; therefore, the training and testing data were combined and used to train the models (giving a total of 137 data points).

Bayes' theorem can be used to estimate the posterior distribution of the weights, $p(\mathbf{w}|\mathbf{y})$, according to:

$$p(\mathbf{w}|\mathbf{y}) \propto p\left(\mathbf{y}|\mathbf{w}, \sigma_{\mathbf{y}}^2\right) p(\mathbf{w}|\sigma_{\mathbf{w}}^2) \tag{14}$$

where $p\left(\mathbf{y}|\mathbf{w}, \sigma_{\mathbf{y}}^2\right)$ is the likelihood function; $p(\mathbf{w}|\sigma_{\mathbf{w}}^2)$ is the prior distribution of the weights; and $\sigma_{\mathbf{y}}^2$ and $\sigma_{\mathbf{w}}^2$ are the variances of the observed data and the network weights, respectively. Based on the above proportionality, the MCMC training method involves a two-step iterative procedure, where ANN weight vectors are sampled from the posterior distribution using the adaptive Metropolis algorithm (Haario et al., 2001) and the likelihood and prior variance hyperparameters are sampled using the Gibbs sampler. It is assumed that the residuals between the observed data y and the model outputs $\hat{y}$ are normally and independently distributed with zero mean and constant variance $\sigma_{\mathbf{y}}^2$; thus, the likelihood function is given by:

$$p\left(\mathbf{y}|\mathbf{w}, \sigma_y^2\right) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \prod_{i=1}^{N} \exp\left\{-\frac{(y_i - \hat{y}_i)^2}{2\sigma_y^2}\right\} \tag{15}$$

A hierarchical prior distribution was chosen to reflect the lack of prior knowledge about the weights and to help prevent overfitting of the data. Given by eq. (14), this prior is the product of four different normal distributions, each with a mean of zero, corresponding to four different groups of weights: the input-hidden layer weights, the hidden layer biases, the hidden-output layer weights and the output layer biases; where $\sigma_{w_g}^2$ and d_g are the variance and dimension of the *g*th weight group, respectively.

$$p(\mathbf{w}|\sigma_{\mathbf{w}}^2) = \prod_{g=1}^{4} (2\pi\sigma_{w_g}^2)^{-d_g/2} \exp\left(\frac{-\sum_{i=1}^{d_g} w_i^2}{2\sigma_{w_g}^2}\right) \tag{16}$$

Both $\sigma_{\mathbf{y}}^2$ and $\sigma_{\mathbf{w}}^2 = \{\sigma_{w_1}^2, \ldots, \sigma_{w_4}^2\}$ in (13) and (14), respectively, are treated as unknown hyperparameters with rather non-informative inverse chi-squared hyperprior distributions, which allows their values to be determined from the data. In terms of the prior distribution, this encourages unnecessary weights to take values close to zero, since a small variance will result in a greater prior probability for such weights. Further details of this method can be found in (Kingston et al., 2005).

Samples from the posterior weight distribution can be used to calculate an *output* distribution for each given input pattern. These outputs must then converted to predictive distributions (i.e. a forecast distribution for each input pattern) by adding back in the Gaussian residuals with zero mean and variance $\sigma_{\mathbf{y}}^2$. As such, total forecast uncertainty is treated in a similar manner to that described in Section 4, where parameter uncertainty is explicitly assessed and all other sources of uncertainty are aggregated into a residual error term. A total of 10,000 weight vectors and corresponding residual variance values were sampled from the posterior weight distribution, $p(\mathbf{w}|\sigma_{\mathbf{w}}^2)$, and the posterior variance distribution, $p(\sigma_{\mathbf{y}}^2|\mathbf{w},\mathbf{y})$, respectively, and subsequently used to calculate forecast distributions.

5.5 Assessment Metrics

In order to assess the quality of the forecasts produced by the various models, and hence whether or not the soil moisture datasets can improve streamflow forecasts, a number of different measures were used. To evaluate the accuracy of the median forecasts (i.e. the medians of the forecast distributions), the root mean squared error (RMSE), Nash-Sutcliffe efficiency (NSE) (Nash and Sutcliffe, 1970) and percent bias were computed for each model.

The RMSE is a measure of overall error between the observed and modelled data and returns an error value with the same units as the data. It is computed based on squared differences as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(O_i - Pmed_i)^2}{N}} \tag{17}$$

where O_i is the *i*th observation, $Pmed_i$ is the *i*th median forecast value (obtained from the predictive distributions) and N is the total number of observations. Squaring the residuals means that larger deviations between the observed and modelled values can have a significantly greater influence on the overall statistic and, as such, the RMSE is biased in favour of peaks and higher magnitude events which tend to result in the poorest predictions (Dawson et al., 2007). The RMSE in non-negative and for a perfect model would be 0.

The NSE statistic is a dimensionless "goodness-of-fit" measure which determines the relative magnitude of residual variance compared to the measured data variance according to:

$$NSE = 1 - \frac{\sum_{i=1}^{N}(O_i - Pmed_i)^2}{\sum_{i=1}^{N}(O_i - \bar{O})^2} \tag{18}$$

where $\bar{O}$ is the mean of observed data. The NSE can range from $-\infty$ in the worst case to 1 for a perfect model. Generally, values between 0 and 1 are viewed as acceptable, whereas values < 0 indicate that the mean observation is a better predictor than the forecast model (Moriasi et al., 2007). The NSE can be useful for benchmarking the performance of the streamflow forecasting models, as a number of authors have considered NSE scores ≥0.9 to be ''very satisfactory'', between 0.8 and 0.9 to be ''fairly good'', and <0.8 to be ''unsatisfactory'' when applied for the assessment of hydrological models (Dawson et al., 2007; Shamseldin, 1997).

Percent bias was calculated according to:

$$bias = \frac{\sum_{i=1}^{N} Pmed_i - \sum_{i=1}^{N} O_i}{\sum_{i=1}^{N} O_i} \times 100\% \quad (19)$$

This metric determines whether forecasts have a general tendency to be too high or too low, where a positive bias indicates over-estimation by the model (the mean of the median forecasts is greater than the mean of the observations), while a negative bias indicates under-estimation (the mean of the median forecasts is less than the mean of the observations).

Since the metrics above were each calculated using the median of the probabilistic forecasts produced by the ANN models, they are reduced to deterministic measures of forecast accuracy. As such, they do not provide any indication of the uncertainty associated with the forecasts and can not be used to test if the forecast distributions are consistent with the observed data. In addition to these metrics, the reliability (α) and precision ($\pi^{(rel)}$) metrics described in Section 4 were used to quantify the reliability and resolution of the probabilistic forecast distributions, respectively. As mentioned in Section 4 two models may yield the same value of α (same reliability of the forecast distributions), but have different uncertainty bounds. While the precision metric, $\pi^{(rel)}$, enables an assessment and comparison of the widths of the uncertainty distributions, it does not indicate whether a model is over-confident (uncertainty has been underestimated resulting in too narrow forecast distributions) or under-confident (uncertainty has been overestimated resulting in too wide forecast distributions). Therefore, a "hit rate", which gives the proportion of observations that fall within the forecast distributions, was used in conjunction with the above metrics to assess the accuracy of the probabilistic forecasts. The hit rate can be calculated by (Demirel et al., 2013):

$$hit\ rate = \frac{hits}{(hits + misses)} \times 100\% \quad (20)$$

where a "hit" occurs when the observation falls within the 95% forecast limits and a "miss" occurs when the observation is not included within the 95% forecast limits. For a reliable forecast model, it would be expected that 95% of the observations should fall within the 95% forecast limits, giving a hit rate of 95%. A hit rate closer to 100% may indicate that forecast uncertainty is overestimated (the model is under-confident), while a hit rate much lower than 95% may either indicate that forecast uncertainty is underestimated (the model is over-confident) or that the model is inaccurate. If the model is in fact inaccurate, this should be evident from the other assessment metrics.

5.6 Results

5.6.1 Input Selection

The results of the GA-ANN IVS are given in Table 9, which lists the inputs selected for each of the five forecast models developed. Despite the relatively large candidate input pools for some of the models (***base, base+POAMA SM***, and ***base+ECV SM*** with between 40 to 57 potential inputs), the GA-ANN has selected a fairly parsimonious set of inputs for all models, with between 7 and 11 inputs. Surprisingly, no previous flows were selected as inputs for the ***base*** model, for which the selected inputs instead consisted of previous rainfall, evaporation, API and groundwater at various locations across the entire Drain M catchment, as well as forecast rainfall from the north eastern part of the catchment (at gauge 26075, see Figure 34). Consequently, no flow inputs were selected for any of the subsequent models which included soil moisture data, since the candidate input pools for these models were restricted to only include the inputs selected for the ***base*** model with the addition of the soil moisture data of interest. It is interesting to note that all six of the inputs selected for representing initial catchment conditions in the ***base*** model were also selected for the models including soil moisture data, with the exception of the $API\_A2390514_{t-2}$ input, which was omitted when GR4J SM data were included (i.e. for the ***base+GR4J SM*** and ***base+all SM*** models). It is also interesting to note that soil moisture data were selected in addition to the ***base*** model inputs in all cases where such data were included in the candidate input pool. These results suggest that the soil moisture data provide relevant information about the rainfall-runoff relationship additional to that characterised by the more traditional predictors of initial catchment wetness, with the exception of API, which becomes redundant with the inclusion of GR4J SM data.

Table 9 Selected inputs for different ANN models

Model	Number of inputs	Selected Inputs
base	7	Rain_A2390519$_t$ Rain_A2390512$_t$ Rain_A2390512$_{t-1}$ Evap_A2390519$_{t-1}$ GW_LI_SMT020$_t$ API_A2390514$_{t-2}$ POAMA_Rain_26075_mean$_{t+1}$
base+ECV SM	9	Rain_A2390519$_t$ Rain_A2390512$_t$ Rain_A2390512$_{t-1}$ Evap_A2390519$_{t-1}$ GW_LI_SMT020$_t$ API_A2390514$_{t-2}$ ECV_SM_A2390515$_t$ ECV_SM_A2391001$_t$ POAMA_Rain_26075_mean$_{t+1}$
base+POAMA SM	9	Rain_A2390519$_t$ Rain_A2390512$_t$ Rain_A2390512$_{t-1}$ Evap_A2390519$_{t-1}$ GW_LI_SMT020$_t$ API_A2390514$_{t-2}$ POAMA_SM_A2390515_mean$_{t+1}$ POAMA_SM_A2391076_max$_{t+1}$ POAMA_Rain_26075_mean$_{t+1}$
base+GR4J SM	9	Rain_A2390519$_t$ Rain_A2390512$_t$ Rain_A2390512$_{t-1}$ Evap_A2390519$_{t-1}$ GW_LI_SMT020$_t$ GR4J_SM_A2390512$_t$ GR4J_SM_A2390512$_{t-1}$ GR4J_SM_A2390512$_{t-2}$ POAMA_Rain_26075_mean$_{t+1}$
base+all SM	11	Rain_A2390519$_t$ Rain_A2390512$_t$ Rain_A2390512$_{t-1}$ Evap_A2390519$_{t-1}$ GW_LI_SMT020$_t$ GR4J_SM_A2390512$_t$ GR4J_SM_A2390512$_{t-1}$ GR4J_SM_A2390512$_{t-2}$ POAMA_SM_A2391076_max$_{t+1}$ ECV_SM_A2391001$_t$ POAMA_Rain_26075_mean$_{t+1}$

The greatest number of inputs were selected for the ***base+all SM*** model (11 compared to 7-9 for the other models), which also had the smallest candidate input pool to select from (a total of 14 candidate inputs). This result implies that the selected inputs are all relevant and necessary for optimal predictive performance, since with a fairly small space to search, the selection of optimal inputs should be relatively easy for the GA-ANN IVS algorithm. The subcatchments from which GR4J SM, ECV SM and POAMA SM inputs were selected for the ***base+all SM*** model are shown in Figure 40, denoted by the purple, pink and yellow shaded areas, respectively. It can be seen that each of the SM inputs are from different areas of the overall Drain M catchment and, thus, provide different information about soil moisture conditions across the catchment. Also shown in Figure 40 is the spatial grid of ECV SM data on which both the ECV SM and POAMA SM inputs are based (POAMA SM uses ECV SM data in the downscaling process). As can be seen, the selected ECV SM and POAMA SM inputs are from subcatchments either closely surrounded by grid points (ECV SM – pink shaded region) or with a single grid point close to the centroid (POAMA SM – yellow shaded region). As such, it would be expected that soil moisture estimates for these subcatchments would be fairly representative of average wetness conditions in these subcatchments, while the soil moisture for some of the other subcatchments may not be quite so well represented. The selected GR4J SM input is from subcatchment A2390512, which makes sense given that the GR4J model used to produce these data was calibrated to flows at A2390512, while other potential GR4J inputs were generated using models calibrated to flows further upstream.

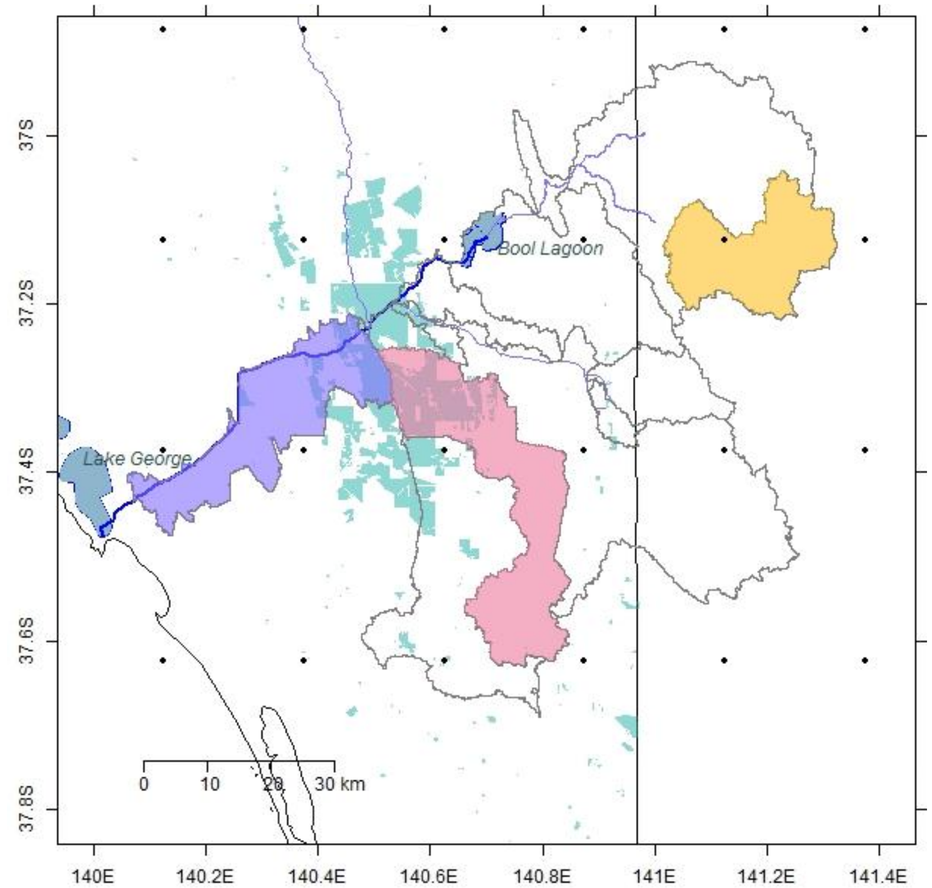


Figure 40 Subcatchments from which GR4J SM (purple), ECV SM (pink) and POAMA SM (yellow) inputs were selected for the *base + all SM* **model**

5.6.2 Model Results

The results of the five streamflow forecasting models developed are presented in Table 10 in terms of the assessment metrics discussed in Section 5.5. As the primary purpose of the models is to accurately and reliably forecast higher flows (i.e. flow volumes sufficient to meet downstream water requirements and allow for diversions to the north) rather than low flows, reliability and precision metrics were calculated based on all flows and for monthly flows above 10GL only. Results are given in Table 10 for both the training and the validation data.

Table 10 Forecast accuracy results for all models

	All flows						Flows > 10GL/month		
	NSE	RMSE	Bias (%)	α	$\pi^{(rel)}$	Hit Rate (%)	α	$\pi^{(rel)}$	Hit Rate (%)
Training									
base	0.87	5624	-2	0.81	1.38	95	0.90	4.29	71
base + GR4J SM	0.89	5296	1	0.74	1.49	95	0.80	4.52	79
base + ECV SM	0.88	5376	0	0.75	1.44	96	0.83	4.45	78
base + POAMA SM	0.88	5372	0	0.75	1.46	95	0.83	4.44	79
base + all SM	0.88	5340	0	0.73	1.41	94	0.80	4.35	74
Validation									
base	0.78	4376	12	0.75	1.42	97	0.80	3.30	88
base + GR4J SM	0.89	3075	10	0.73	1.48	100	0.65	3.42	100
base + ECV SM	0.81	4027	13	0.75	1.51	97	0.75	3.36	88
base + POAMA SM	0.76	4530	18	0.76	1.53	100	0.74	3.48	100
base + all SM	0.88	3262	10	0.75	1.45	100	0.79	3.40	100

Overall, all of the models performed well on both sets of data, with most resulting NSE values > 0.8. The performance results of the models on the training data were very similar, with no model being obviously better than another according to the different metrics. However, when considering performance on the validation data, there was a noticeable improvement in median forecast accuracy with the inclusion of GR4J SM inputs (***base+GR4J SM*** and ***base+all SM*** models), where an increase in NSE and reduction in RMSE (by 25-30%) and percent bias were seen. There was also a slight improvement gained with the addition of ECV SM inputs; however, the median performance results of the ***base*** and ***base+ECV SM*** models are similar enough that any differences are considered fairly insignificant. The inclusion of POAMA SM resulted in a slight reduction in median forecast performance, as seen by the RMSE, NSE and percent bias results; however, the predictive distribution of this model was the most reliable and precise according to the α and $\pi^{(rel)}$ metrics, respectively, when evaluated based on all flows. In fact, the addition of any soil moisture information resulted in more precise forecast distributions (more confident models) on both the training and validation data when compared to the ***base*** model.

When considering all flows, all of the models returned hit rates of > 95% for the validation data, which may indicate that forecast uncertainty has been overestimated by each of the models, particularly those with 100% hit rates. However, there were only 30 values in the validation dataset and the high hit rates may simply be a result of the small sample size. Given that the hit rates for the training data were all around 95% when based on all flows, and the predictive distributions for the validation data are at least as precise as those for the training data (as seen by the $\pi^{(rel)}$ values calculated based on all flows), this seems to be the case. However, when considering the high flows only, hit rates for the training data were significantly less than 95%. These results, together with the high α and $\pi^{(rel)}$ values associated with the high flow training data forecasts, indicate that forecast uncertainty has been underestimated for flows above 10GL/month. However, this was not apparent for the validation high flows for which the uncertainty bounds were wider (as seen by the lower $\pi^{(rel)}$ values in comparison to the $\pi^{(rel)}$ values for the training data) and hit rates of 100% were still returned by three of the models. Nevertheless, it can be seen in Table 10 that there was a significant increase in the precision of the forecast distributions ($\pi^{(rel)}$) for both the training and validation data when only flows above 10GL/month were considered, indicating that the relative uncertainty associated with lower flows (< 10GL/month) is significantly greater than that associated with the higher flows. This can also be seen in Figure 41, which gives a visual comparison of the reliability, α, and precision, $\pi^{(rel)}$, metrics across all models when computed based on all flows or high flows only. As can also be seen in this figure, the inclusion of soil moisture inputs resulted in a reduction in α when applied to the training data, for all flows and high flows >10GL/month (unlike precision, which increased). This was also seen when the models were applied to the validation high flows, while the α values remained very similar across all models when all flows were considered. In comparison to the ***base*** model, the most significant reduction in forecast reliability on the validation data (particularly for high flows) was seen when GR4J SM inputs only were included in the ANN model, despite the fact that this model achieved the best median performance on the validation data. However, only the ***base*** and ***base+ECV SM*** models resulted in hit rates of less than 95% on the high flows.

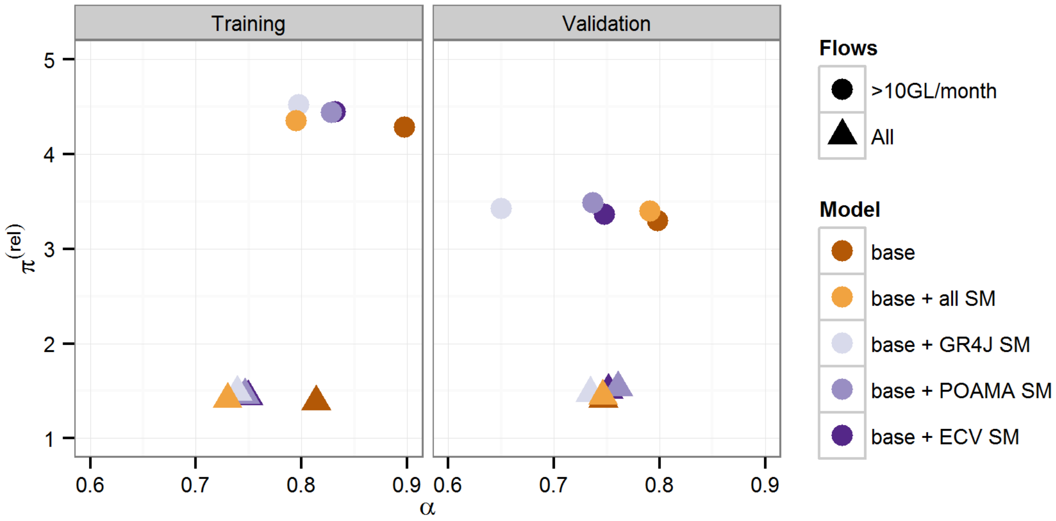


Figure 41 Reliability (α) and precision ($\pi^{(rel)}$) for the 5 different forecast models

Overall, the ***base+all SM*** model was considered to give the best forecasts, with good median forecast accuracy and relatively reliable and precise forecast distributions, particularly for flows above 10GL/month, which are of primary interest when trying to maximise the use of such flows. Shown in Figure 42 is a time series plot of the median forecasts and 95% forecast limits resulting from this model in comparison to the observed flow at A2390512 and 95% limits of the historical flow distributions by month. As can be seen, the model gives a significantly better forecast of 1-month ahead flows than the historical flow distributions. Of the models that only considered a single soil moisture dataset, the ***base+GR4J SM*** model achieved the biggest improvement in streamflow forecasts over the ***base*** model, while the improvements gained from the inclusion of ECV SM and POAMA SM inputs were less apparent. However, the inclusion of GR4J SM data resulted in a reduction in the reliability of the forecast distributions, which was corrected when the other two sources of soil moisture data were also included.

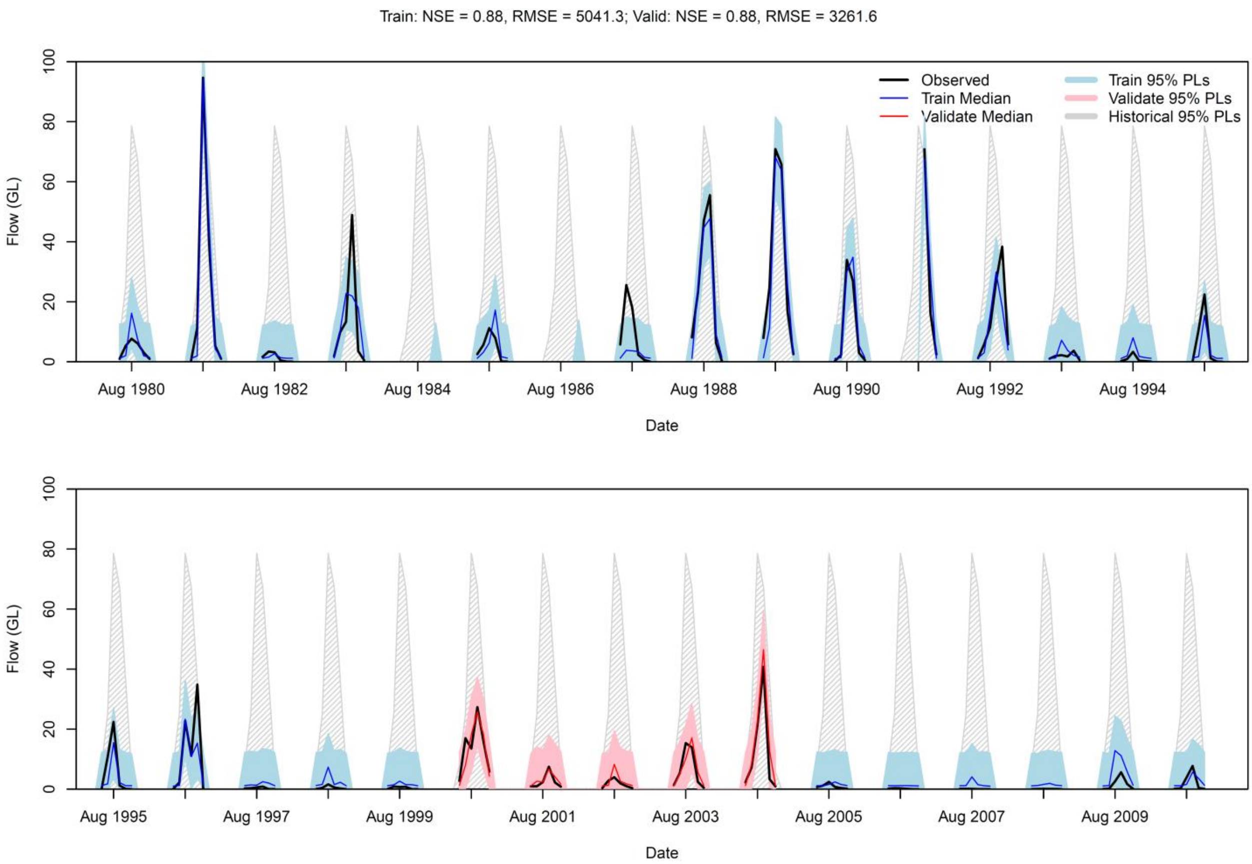


Figure 42 Median forecasts and 95% forecast limits resulting from the *base+all SM* model in comparison to observed flow at A2390512 and historical monthly flow distributions.

5.7 Conclusions

In this section, the use of the GR4J SM, ECV SM and POAMA SM datasets as direct inputs to an ANN forecast model, developed to forecast flow at gauge A2390512 one month in advance, was explored in order to identify whether the information content of these datasets is sufficient to improve statistical streamflow forecasts either individually or in combination. Overall, it was found that the inclusion of all three datasets gave the best improvement in forecast accuracy and reliability over a ***base*** model, which only included more traditional inputs for representing antecedent catchment conditions and did not include soil moisture data directly. Nevertheless, it was also found that with the inclusion of GR4J generated soil moisture data alone (***base+GR4J SM***), almost equivalent forecast accuracy and reliability to that of the ***base+all SM*** model was achieved. However, as discussed in Section 3, the performance of the GR4J model may not always be adequate for representing catchment processes, particularly in cases where there is a change in streamflow due to changes in groundwater level, and as such, the utility of GR4J SM data warrants further investigation on additional catchments. Given the slightly increased precision of the probabilistic forecast distributions achieved with the inclusion of ECV SM and POAMA SM data, which did not necessarily come at the expense of forecast reliability, it is considered worthwhile to at least include soil moisture data obtained from large, readily available global datasets (such as POAMA SM and ECV SM) as potential model inputs when developing statistical streamflow forecasting models. It is not recommended, however, that soil moisture data be included at the expense of more traditional types of inputs used for characterising initial catchment conditions, since the results of the input variable selection performed found that such inputs, including previous rainfall, evaporation and groundwater, are also necessary for making accurate streamflow forecasts. Rather, all potentially relevant inputs should be considered, with a rigorous IVS method applied to select those that are most relevant to the problem at hand.

6 Comparison of Statistical and Conceptual Models

In Sections 3 to 5, different lumped conceptual and statistical models (as well as hybrid models) were developed to forecast streamflow in the Drain M catchment one month in advance. The purpose of this section is to compare the outputs of these models in order to identify the most suitable approach for making forecasts for this system.

6.1 Lumped conceptual Models

The lumped conceptual rainfall-runoff (CRR) model used for comparison to the statistical models, and to produce the GR4J SM data used as an input to the statistical models, was the standard GR4J model outlined in Section 3. Parameter uncertainty was estimated using the DREAM approach, and residual error models calibrated as outlined in Section 4. This included a sum of squared errors based likelihood function, assimilation of previous flow data into the routing store to correct for model bias, and a log transformation applied to the data prior to fitting the residual error model (see Section 4). As noted in Section 4, the calibration period used was the previous 10 years of data, and the calibration was repeated at the start of the flow season each year (assumed to be May 1).

For direct comparison with the results from the statistical models, the CRR results in this section represent the total flow at each location, as opposed to previous results presented for the CRRs, which only considered the flow generated within each the sub-catchment. There is no difference between the two cases for the A2390519 catchment, but for the A2390514 catchment, the releases from Bool Lagoon must be added to the flow generated within the catchment, and this total flow must be added with the flow generated within the most downstream catchment for the A2390512 catchment. Releases from Bool Lagoon were not explicitly modelled for use with the CRR models, however a Source model has been developed for this purpose within this project (See Section 7). For the purposes of comparison, the total observed release has been adopted, based on the assumption that the total planned release volume likely to be known one month in advance. This is only relevant for one month in the 5 year validation period, where 8.3 GL was released from Bool Lagoon in September 2004, and the other months there was no flow release recorded from Bool lagoon.

6.1.1 Climate Input Comparison

Rainfall and evapotranspiration time series are required inputs to use the CRRs for forecasting purposes. Three options to provide this data have been considered:

1) the POAMA climate forecast for the month, using the analogue downscaling approach for each of the 33 POAMA models in the ensemble,
2) the climate data for each month over the past 30 years corresponding forecast month (i.e an uninformed forecast using the past 30 years of data for June to forecast for June), and
3) the actual climate data for the forecast month.

The actual data input does not represent a true forecast, as data from the “future” is used, but this input does allow the performance of the CRRs to be assessed, and represents an assumption of a perfect climate forecast.

The median monthly rainfall for the first two forecast rainfall inputs has been compared to the third observed rainfall input for the model validation period. This can be seen in Table 11, where only the months in the forecast period have been considered each year (i.e. June – November). It can be seen that based on NSE and RMSE metrics, the median of the POAMA model ensemble does not improve the rainfall forecast over the historical average. Plots of these data for the model validation period from 2000-2004 can also be seen in Figure 43.

Table 11 Accuracy of long term average (1960-2014) and median of ensemble of POAMA model compared to the observed rainfall for the winter (June-Nov) period from 1980-2011.

Catchment	Rainfall	NSE	RMSE (mm/month)
A2390512	Average	0.28	27.45
	POAMA	0.16	29.74
A2390514	Average	0.4	29.03
	POAMA	0.05	36.47
A2390519	Average	0.33	29.05
	POAMA	0.22	31.18

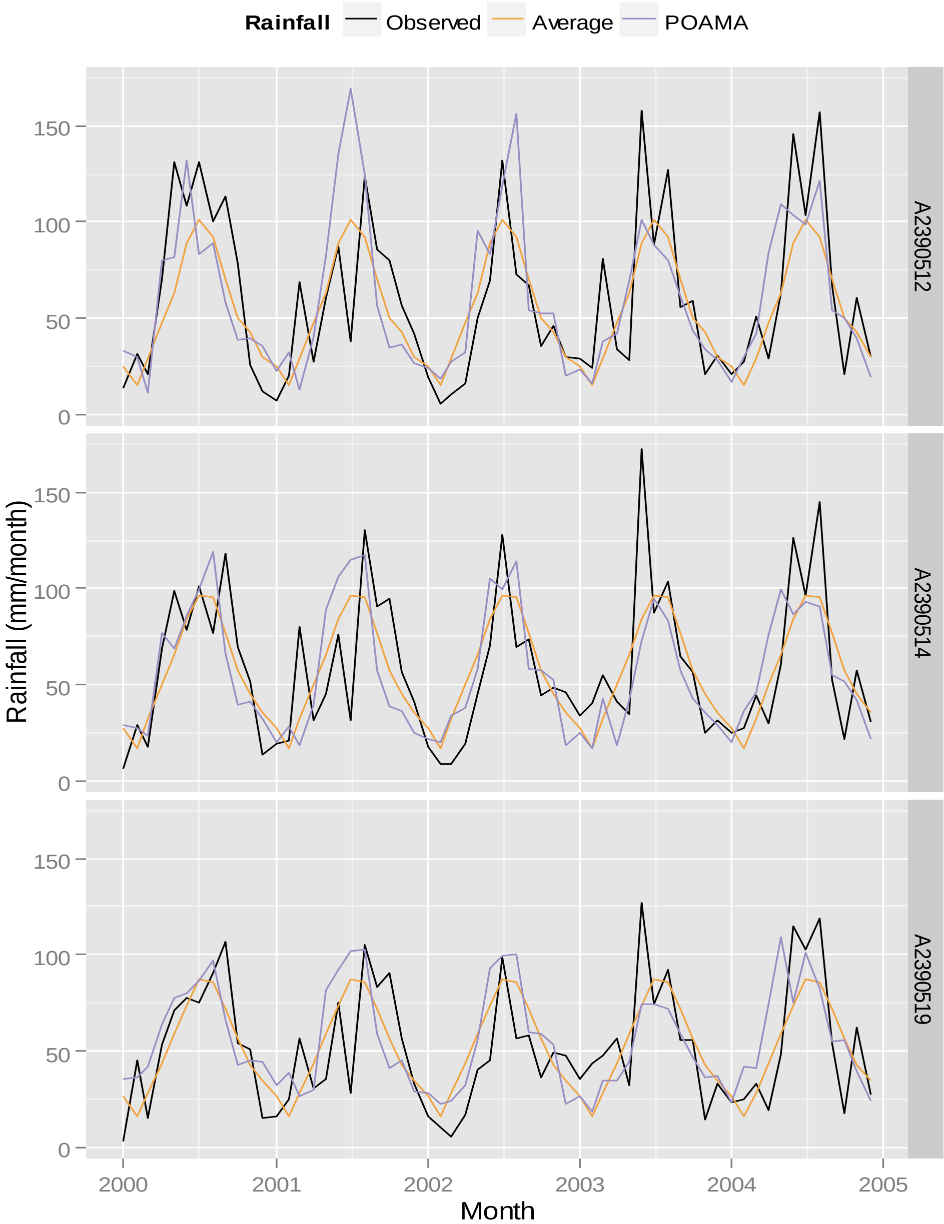


Figure 43 Plot of Observed, long term average (1970 - 1999) and median of POAMA ensemble for the model validation period

6.1.2 CRR Flow Forecasting Results

Results for forecasting the flow in Drain M one month in advance using the CRRs are presented in Table 12. The reliability, resolution and hit rate metrics represent different components of the distribution of flow forecast (see Section 5.5). The NSE, percent bias and RMSE metrics were calculated from the median modelled flow over the ensemble of outputs for the CRR.

Table 12 Forecast results from the conceptual rainfall runoff models for the validation period of June 2000- Nov 2004. For each of the three catchments, three rainfall inputs have been considered. The reliability, resolution and hit rate represent the accuracy of the forecast distribution of monthly flows, while the remaining metrics are calculated from the median climate forecast and best parameter set.

Rainfall	NSE	RMSE (ML/month)	Bias (%)	α	$\pi^{(rel)}$	Hit Rate (%)
			A2390512			
POAMA	-0.16	10145	38	0.85	1.32	73
Historical	0.28	14784	-13	0.68	4.67	43
Actual	0.92	2661	8	0.85	1.19	87
			A2390514			
POAMA	-0.38	5642	28	0.75	0.91	87
Historical	-0.11	3254	56	0.64	1.63	60
Actual	0.83	1955	14	0.69	0.88	100
			A2390519			
POAMA	-2.08	4987	53	0.72	0.86	97
Historical	-0.88	6521	87	0.59	0.86	97
Actual	-0.98	3995	150	0.61	2.35	50

6.1.3 When considering the predictive distribution, it can be seen from CRR Flow Forecasting Results

Results for forecasting the flow in Drain M one month in advance using the CRRs are presented in Table 12. The reliability, resolution and hit rate metrics represent different components of the distribution of flow forecast (see Section 5.5). The NSE, percent bias and RMSE metrics were calculated from the median modelled flow over the ensemble of outputs for the CRR.

Table 12 that the POAMA forecast was informative, with the predictive distribution more reliable (but less precise) compared to the historical rainfall input. Across all runs it can be seen that there is a strong correlation between the resolution of the predictive distribution and the Hit Rate, where as the predictive distribution becomes wider (lower resolution values), the number of observed values that fall within the 95th percentile confidence bounds (Hit Rate) also increases.

The results comparing the best CRR for each catchment and median simulated flow produced from the range of rainfall inputs (only one rainfall input for the Actual rainfall case), indicate that two of the models perform well when using observed rainfall as the input, with the NSE > 0.8 and Bias < 15% for the A2390512 and A2390514 catchments. It can be seen from Table 12 that in general the median of the historical range of rainfall for that month provided better results than the median of the ensemble of POAMA forecast rainfall, most likely due to the better performance of the median monthly rainfall when compared to the observed rainfall, as seen in Table 11.

Poor performance can be seen for the A2390519 catchment, with negative NSE value and a bias value indicating an over prediction of the total volume in the validation period of 150%. This result is expected to be due to the change in the catchment rainfall runoff relationship, seen in Figure 7. If it is assumed there was a change in the slope between cumulative rainfall and cumulative runoff around 1995, then this change occurred in the middle of the 10 year calibration period (1990-1999) for the model used to simulate the 2000 validation period. To further test this case, the flow simulated using the maximum likelihood model parameter values for each year for both 10 year and 5 year calibration periods, as well as the observed flow, is presented in Figure 44. It can be seen that both the models substantially overestimated the observed flow in 2000, possibly due to this calibration to a wetter runoff regime while they were used to simulate a drier regime. By 2004 it can be seen that the model with the 10 year model calibration period (i.e. 1994-2003) still overestimated the observed flow substantially, but the model that was calibrated to only more recent 5 year period (1999-2003) provided a more accurate representation of the observed flow. As such, improved model performance may be able to be achieved for this catchment on this period if a different, more representative, calibration period was adopted.

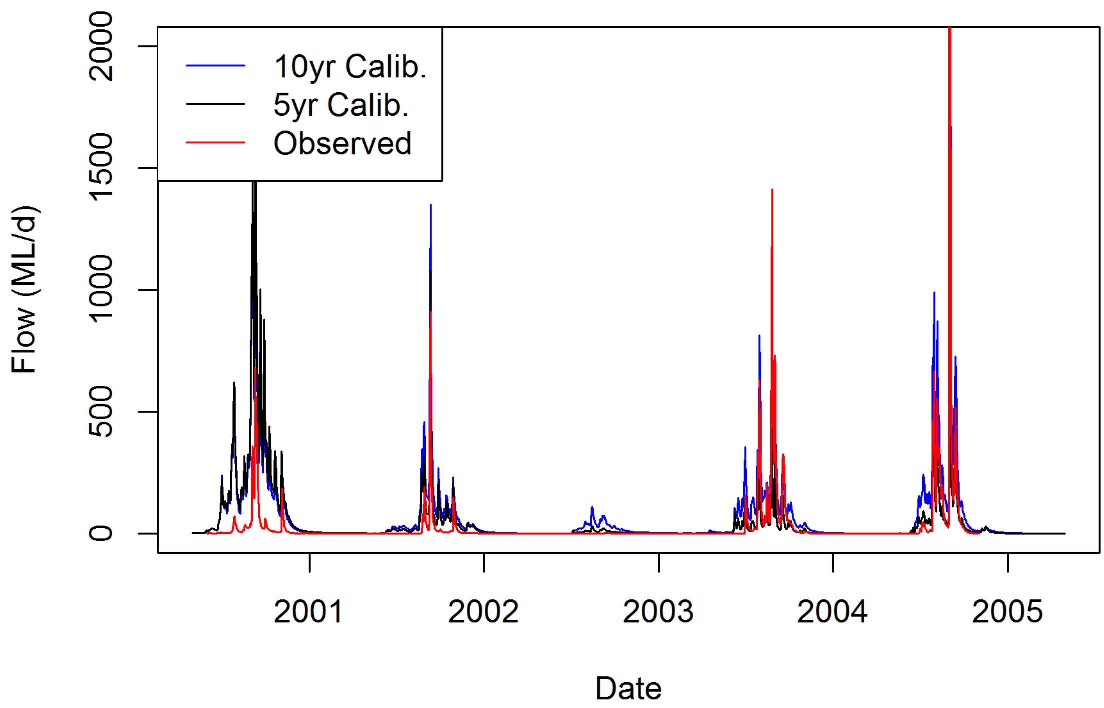


Figure 44 Comparison between observed and modelled flow for the A2390519 catchment and validation period. Two different calibration periods are presented, and it can be seen the shorter calibration period provides a more accurate representation of the observed flow, particularly toward the end of the period.

6.2 Statistical Models

In Section 5 it was identified that the hybrid models, which used GR4J generated soil moisture data as inputs (i.e. ***base+GR4J SM*** and ***base+all SM***), resulted in the most accurate and reliable forecasts of the statistical models considered when used to forecast flow at gauge A2390512, at the bottom of the Drain M system. However, given the differing characteristics of the three Drain M catchments (see Figure 3) in terms of size (details given in Table 1), topography, detention storage and landuse, it should not be assumed that the addition of GR4J soil moisture data would similarly improve the performance of the ANN models when applied to forecast flows for catchments A2390514 and A2390519, which, as noted in Section 2.1, had a more notable change in rainfall-runoff relationship as a result of landuse change when compared to A2390512 (see Figure 7). Furthermore, as seen in Section 6.1, when using actual rainfall data as an input (since the GR4J soil moisture inputs were generated using actual rainfall data), the GR4J model had varying performance when applied to the three catchments, with forecasts generated for catchments A2390514 and A2390519 being less accurate and reliable than those for catchment A2390512.

For this comparison, four ANN models were compared in their abilities to accurately and reliably forecast streamflow at each of the three gauges along the Drain M system (12 models in total):

- a '***base***' model with no soil moisture inputs;
- a '***base+GR4J SM***' model: ***base*** model + GR4J simulated soil moisture inputs;

- a '***base+ECV SM+POAMA SM***' model: ***base*** model + remotely sensed ECV soil moisture inputs + POAMA forecast soil moisture inputs; and
- a '***base+all SM***' which includes all soil moisture data sets as inputs.

Only a combined ***base+ECV SM+POAMA SM*** model was considered, rather than considering the utility of the ECV SM and POAMA SM inputs individually, as it was assumed that the IVS algorithm would only select either ECV SM or POAMA SM inputs should the combination of these inputs not result in better model performance than either of the soil moisture inputs individually.

Consideration of the above four ANN models at each location allows for a comparison between: 1) the ANN method without any soil moisture information and 2) ANN models which include soil moisture information, where this information may or may not include information generated from the GR4J model. Comparison between the latter three models enables an assessment of any improvements in model performance above the ***base*** model and which SM inputs such improvements may be attributed. It also allows for an assessment of whether it is better not to include the GR4J SM information for catchments where the underlying GR4J model may not perform so well or whether any improvements in model performance can still be attributed to the inclusion of GR4J SM in such cases.

It should be noted that the ***base+ECV SM+POAMA SM*** model was the only additional model developed for the A2390512 catchment, while all other results presented for this catchment were obtained using the models developed in Section 5. As such, results presented for the ***base***, ***base+GR4J SM***, and ***base+all SM*** models for catchment A2390512 will be the same as those presented in Section 5 and only nine additional models were developed for the purposes of the comparison presented in this section. Data used for developing the ANN models were the same as those discussed in Section 5.3.

6.2.1 Input Variable Selection

The GA-ANN IVS method outlined in Section 5.4.1 was used to identify the most suitable inputs to the nine additional ANN models developed. Similar to the approach discussed in Section 5.4.1, the candidate input pools for each model were restricted such that they could only include inputs that had already been selected as optimal for the ***base*** model, with the addition of lagged values of the soil moisture data of interest (or minimum, mean and maximum values in the case of the POAMA forecast soil moisture data), while the candidate input pool for the ***base+all SM*** model was further restricted to only include inputs already found to be optimal for the previous three models. Furthermore, only input variables recorded within or upstream of the catchment under consideration were included as potential inputs (e.g. rainfall data for catchment A2390512 was not included in the candidate input pools for any of the models developed for catchments A2390514 or A2390519). The candidate inputs available to be selected for the statistical models are given in Appendix A for each of the cases considered in this section.

The results of the GA-ANN IVS are presented in Table 13 - 15 for catchments A2390512, A2390514 and A2390519, respectively (with previous results for A2390512 presented in Table 9). As can be seen, similar to the A2390512 ***base*** model (see Section 5.6.1), no flow inputs were selected for the A2390514 or A2390519 models, with flow forecasts primarily characterised by rainfall, evaporation and groundwater (where available) inputs in the respective ***base*** models. For catchment A2390514, all of the ***base*** model inputs were selected for the subsequent models which included soil moisture information, with the exception of the Evap_A2390519$_t$ input, which was omitted from all subsequent models. While it is possible that this input became redundant with the inclusion of additional soil moisture information, it is also possible that the GA-ANN solution for the ***base*** model was not entirely optimal, with an additional input being selected than was necessary. For catchment A2390519, on the other hand, it is more likely that input API_A2390519$_{t-1}$, which was selected for the ***base*** model but omitted from all subsequent models, did become redundant with the addition of soil moisture inputs, since the candidate input pool for this catchment's ***base*** model was much smaller than that for the ***base*** model of catchment A2390514 (18 inputs in comparison to 39 inputs), making it easier for the GA-ANN IVS algorithm to find the optimal solution. Nevertheless, it can not be said that any of the soil moisture data can be used to replace any particular input or subset of inputs, as this result was not consistent across the different catchments. It is apparent, however, that rainfall, evaporation and groundwater (where available) data are the minimum input requirements for representing catchment conditions, with each of the soil moisture data sets providing additional information about the soil moisture states (since these inputs were always selected when included in the candidate input pool).

In terms of the GR4J SM data, it can be seen in Table 13 - 15, together with Table 9, that GR4J SM data were selected in every case where these data were included as candidate inputs, suggesting that the addition of this input information always resulted in improved model accuracy (since the GA-ANN algorithm was used to select inputs based on predictive performance). However, it can be seen in Table 14 that, for catchment A2390514, the GR4J SM inputs selected were those generated for the upstream catchment A2390519. It is likely that this result has limited physical interpretation, since the influence of the upstream catchment on flow at A2390514 is largely related to releases from Bool Lagoon, which happen very infrequently (as seen in the time series of flow at gauge A2390541 in Figure 6). However, it is possible that the catchment processes, in particular, those related to soil moisture, are better captured by the GR4J model developed for catchment A2390519, given the relatively poor performance of the GR4J model on the A2390514 catchment (e.g. Figure 24).

Considering the ***base+all SM*** models for each catchment, the number of selected inputs in each case was similar, with 11, 12 and 9 inputs selected for catchments A2390512, A2390514 and A2390519, respectively. However, of these selected inputs, the proportions of 'traditional' inputs used for representing initial catchment conditions varied. For catchments A2390512 and A2390514, around half of the selected inputs were the more traditional predictors of initial catchment wetness (e.g. rainfall, evaporation, groundwater), while only two such inputs were selected for catchment A2390519 (~22% of the total selected inputs). Unlike catchments A2390512 and A2390514, there were no available groundwater data for catchment A2390519, nor were there any upstream hydrological or climate data that were likely to be relevant for forecasting flows in this catchment. As such, additional soil moisture inputs were selected to provide information about the upcoming flows in this catchment. This highlights one of the advantages of ANN models, and statistical models in general, which is that these models are able to take advantage of whatever data are available, rather than relying on a specific set of inputs as more physically-based models do.

Table 13 Selected inputs for catchment A2390512

Model	Number of inputs	Selected Inputs
base+ECV SM+POAMA SM	10	Rain_A2390519$_t$ Rain_A2390512$_t$ Rain_A2390512$_{t-1}$ Evap_A2390519$_{t-1}$ GW_LI_SMT020$_t$ API_A2390514$_{t-2}$ ECV_SM_A2390515$_t$ ECV_SM_A2391001$_t$ POAMA_SM_A2390515_mean$_{t+1}$ POAMA_Rain_26075_mean$_{t+1}$

Table 14 Selected inputs for catchment A2390514

Model	Number of inputs	Selected Inputs
base	8	Rain_A2390514$_{t-1}$ Rain_A2390519$_{t}$ Rain_A2390519$_{t-2}$ Evap_A2390514$_{t}$ Evap_A2390519$_{t}$ API_A2390519$_{t-2}$ GW_LI_CMM079$_{t-1}$ POAMA_Rain_26075_mean$_{t+1}$
base+ECV SM+POAMA SM	10	Rain_A2390514$_{t-1}$ Rain_A2390519$_{t}$ Rain_A2390519$_{t-2}$ Evap_A2390514$_{t}$ API_A2390519$_{t-2}$ GW_LI_CMM079$_{t-1}$ ECV_SM_A2390514$_{t-2}$ POAMA_SM_A2390519_min$_{t+1}$ POAMA_SM_A2391076_mean$_{t+1}$ POAMA_Rain_26075_mean$_{t+1}$
base+GR4J SM	9	Rain_A2390514$_{t-1}$ Rain_A2390519$_{t}$ Rain_A2390519$_{t-2}$ Evap_A2390514$_{t}$ API_A2390519$_{t-2}$ GW_LI_CMM079$_{t-1}$ GR4J_SM_A2390519$_{t}$ GR4J_SM_A2390519$_{t-1}$ POAMA_Rain_26075_mean$_{t+1}$
base+all SM	12	Rain_A2390514$_{t-1}$ Rain_A2390519$_{t}$ Rain_A2390519$_{t-2}$ Evap_A2390514$_{t}$ API_A2390519$_{t-2}$ GW_LI_CMM079$_{t-1}$ GR4J_SM_A2390519$_{t}$ GR4J_SM_A2390519$_{t-1}$ ECV_SM_A2390514$_{t-2}$ POAMA_SM_A2390519_min$_{t+1}$ POAMA_SM_A2391076_mean$_{t+1}$ POAMA_Rain_26075_mean$_{t+1}$

Table 15 Selected Inputs for Catchment A2390519

Model	Number of inputs	Selected Inputs
base	5	$Rain\_A2390519_t$ $Evap\_A2390519_t$ $Evap\_A2390519_{t-1}$ $API\_A2390519_{t-1}$ $POAMA\_Rain\_26082\_mean_{t+1}$
base+ECV SM+POAMA SM	9	$Rain\_A2390519_t$ $Evap\_A2390519_t$ $Evap\_A2390519_{t-1}$ $ECV\_SM\_A2390519_t$ $ECV\_SM\_A2390519_{t-1}$ $ECV\_SM\_A2391076_t$ $ECV\_SM\_A2391076_{t-1}$ $POAMA\_SM\_A2390519\_max_{t+1}$ $POAMA\_Rain\_26082\_mean_{t+1}$
base+GR4J SM	5	$Rain\_A2390519_t$ $Evap\_A2390519_t$ $Evap\_A2390519_{t-1}$ $GR4J\_SM\_A2390519_t$ $POAMA\_Rain\_26082\_mean_{t+1}$
base+all SM	9	$Rain\_A2390519_t$ $Evap\_A2390519_t$ $GR4J\_SM\_A2390519_t$ $ECV\_SM\_A2390519_t$ $ECV\_SM\_A2390519_{t-1}$ $ECV\_SM\_A2391076_t$ $ECV\_SM\_A2391076_{t-1}$ $POAMA\_SM\_A2390519\_max_{t+1}$ $POAMA\_Rain\_26082\_mean_{t+1}$

6.2.2 ANN Model Development

As discussed in Section 5.4.2, the ANN model development procedure required that the available data be divided into training, testing and validation subsets. For each of the 12 ANN models (as well as the conceptual models) compared in this section, the five years between 2000 to 2004 (inclusive) were used for independent validation of the models, while the remaining data between 1980 and 2010 were used for training and testing. The validation data were selected such that a mixture of both relatively high and low flows were present, but also to exclude the extremes of the data such that these could be used for calibration. Furthermore, the validation period represents the “current” rainfall-runoff relationship, which changed around the late 1990s after the establishment of large areas of plantation forestry (see Figure 7). As such, performance results on these data should give a better representation of how the models may perform if used for operational purposes. In order to create the training and testing data sets, the remaining data were repeatedly randomly divided into subsets with proportions of 80% and 20%, respectively, until the statistics of the training and testing portions of the data were most similar. After accounting for the appropriate lags of the input and output variables, the number of monthly winter and spring data available for training, testing and validation were 125, 31 and 30, respectively, for each of the three catchments.

The MCMC Bayesian training approach discussed in Section 5.4.2 was again used to calibrate the models (using combined training and testing data – 156 data points) and provide estimates of forecast uncertainty. The forecast accuracy results for each model can be seen in Table 16, according to the assessment metrics discussed in Section 5.5. These results are given for the training data only, while the validation results are presented in the following section in comparison to the GR4J conceptual model results.

Table 16 ANN Training Dataset Results

	NSE	RMSE	Bias (%)	α	$\pi^{(rel)}$	Hit Rate (%)
			A2390512			
base	0.87	5624	-2	0.81	1.38	95
base+GR4J SM	0.89	5296	1	0.74	1.49	95
base+ECV SM+POAMA SM	0.89	5248	0	0.80	1.43	95
base+all SM	0.88	5340	0	0.73	1.41	94
			A2390514			
base	0.74	6433	-1	0.60	1.11	95
base+GR4J SM	0.79	5759	-2	0.60	1.21	94
base+ECV SM+POAMA SM	0.78	5489	0	0.56	1.18	95
base+all SM	0.84	4762	-1	0.59	1.30	95
			A2390519			
base	0.47	5306	-1	0.72	0.84	96
base+GR4J SM	0.51	5109	-1	0.72	0.88	96
base+ECV SM+POAMA SM	0.62	4320	1	0.68	0.95	92
base+all SM	0.61	4352	-1	0.69	0.96	91

Considering the results in Table 16 alone, it is apparent that there is very limited value in including GR4J SM inputs, or any soil moisture inputs, for forecasting flows at A2390512, as the accuracy and reliability results are similar across all models, including the ***base*** model with no soil moisture inputs. There is, however, a slight improvement in median accuracy and resolution of the forecast distribution when ECV SM and POAMA SM inputs are included. For catchment A2390514, there is a marginally more obvious improvement in median forecast accuracy when some source of soil moisture information is included in the model (although not necessarily GR4J SM), while the reliabilities and resolutions of the forecast distributions remain similar across all models (there is some improvement in resolution when soil moisture is included). Finally, for catchment A2390519, the value of including soil moisture inputs is most apparent, where median forecast accuracy is up to ~20% better according to the RMSE when soil moisture inputs are included in the model. For this catchment, the greatest improvement in median performance was achieved with the addition of ECV SM and POAMA SM inputs, while the addition of GR4J SM inputs appears to offer little value. However, all models perform relatively poorly on this catchment with all NSE values <0.7. As mentioned previously, there were no groundwater data available for catchment A2390519, which seems necessary for representing changes in streamflow due to changes in groundwater level.

For all of the 12 models considered, hit rates of around 95% were obtained indicating that forecast uncertainty associated with the training data appropriately accounted for the historical flows. However, in order to assess whether any of the models were weak in any particular area (i.e unable to appropriately forecast a particular range of flow magnitudes), scatter plots of the median forecasts and 95% prediction limits generated using the best performing models for each catchment were assessed. These plots are shown in Figures Figure 45 - Figure 47, where the best performing models were considered to be the ***base+GR4J SM***, ***base+all SM*** and ***base+ECV SM+POAMA SM*** models for catchments A2390512, A2390514 and A2390519, respectively, based on all of the metrics presented in Table 16. As can be seen in Figure 45, for A2390512 flows, the upper and lower 95% forecast limits (denoted by blue and red points, respectively) generally encompass the 1:1 line (red dashed line). However, there are several observed flows with magnitudes between 25-50 GL/month that have been underpredicted by the model, as seen by the upper and lower forecast limits which both plot below the 1:1 line. Nevertheless, overall, the median forecasts plot reasonably well along the 1:1 line across all magnitudes of flow. For A2390514 flows, on the other hand, the model appears to be insufficient for forecasting and quantifying the uncertainty associated with the extreme flows, with the maximum observed flows not being accounted for within the forecast uncertainty limits slope of the median forecasts tapering off with higher flows, as seen in Figure 46. On the contrary, it is the mid-range flows between 15-30GL/month that are not forecast well by the best model for A2390519, as seen in Figure 47. Using this model, several of the low flows have been overestimated, while many of the mid-range flows have been underestimated and are not accounted for within the estimated forecast uncertainty. On the other hand, the extreme high flows appear to have been forecast relatively well.

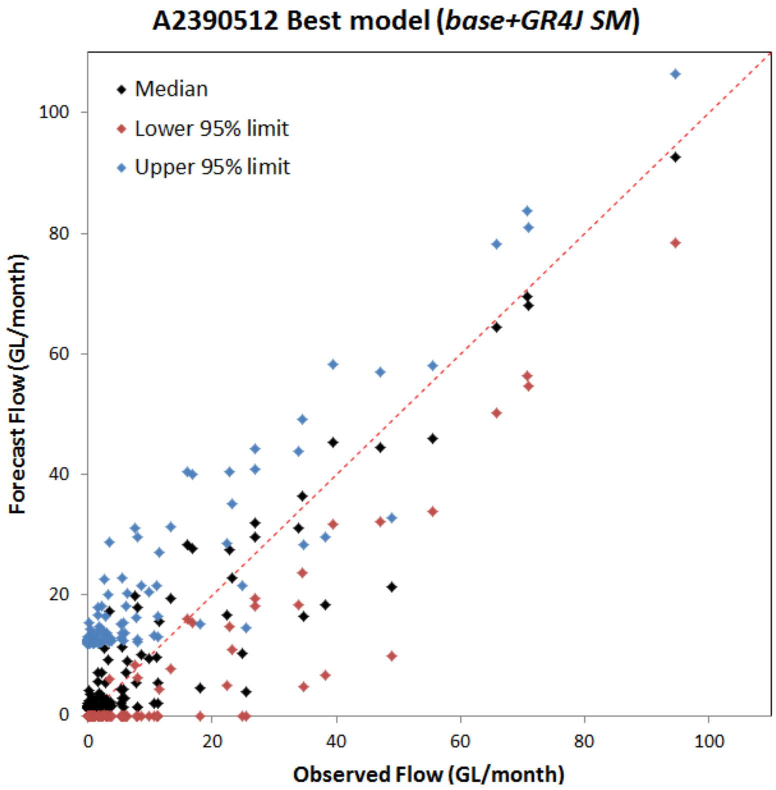


Figure 45 Scatterplot of median forecast flows and 95% uncertainty limits versus observed monthly flows for catchment A2390512 best model (*base+GR4J SM*).

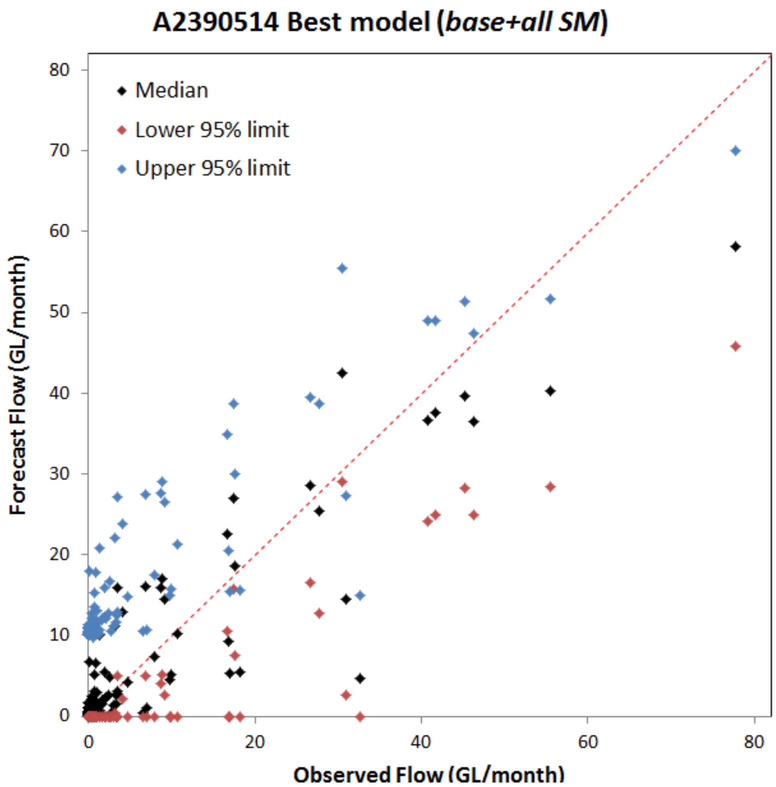


Figure 46 Scatterplot of median forecast flows and 95% uncertainty limits versus observed monthly flows for catchment A2390514 best model (*base+all SM*).

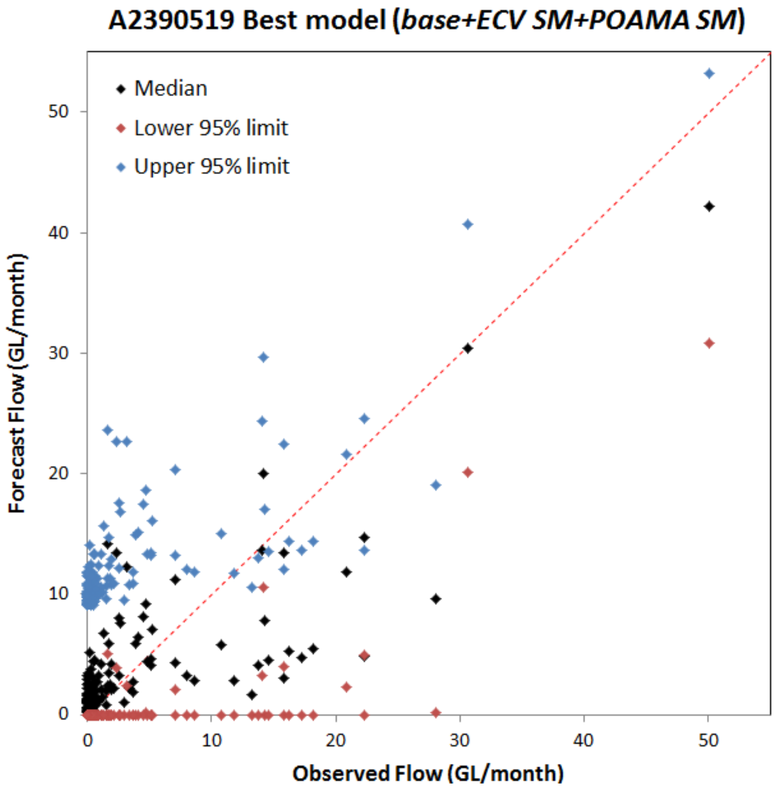


Figure 47 Scatterplot of median forecast flows and 95% uncertainty limits versus observed monthly flows for catchment A2390519 best model (*base+ECV SM+POAMA SM*).

6.3 Results of Model Comparison

In the previous section, the comparison of statistical models with and without different sources of soil moisture data when applied to the training data showed that, for each Drain M catchment, some improvement could be gained from the inclusion of soil moisture inputs; however, it was not immediately obvious whether the inclusion of GR4J SM data could improve model performance over the different catchments where the GR4J model may be more appropriate than others. In this section, the accuracy and reliability results of the ANN models are compared when applied to the validation data. Furthermore, these models are also compared to the GR4J model which utilised POAMA rainfall forecasts presented in Section 6.1, in order to make an assessment of which models are most suited to providing streamflow forecasts in each of the Drain M catchments. The GR4J model with POAMA rainfall input best represents a true forecast model, as it is based on forecasts of the upcoming climate.

The results of these models for the validation period are presented in Table 17 in terms of the metrics discussed in Section 5.5. When considering the gain achieved by including the GR4J SM input in the statistical models, it can be seen that the addition of this input resulted in improved median forecasts in comparison to the ***base*** model for all catchments, as indicated by the NSE, RMSE and percent bias metrics. For catchments A2390512 and A2390514, the inclusion of GR4J SM data achieved the largest improvement in streamflow forecasts over the ***base*** model according to the deterministic metrics; however, for catchment A2390519, the addition of ECV SM and POAMA SM resulted in a greater improvement in median forecast accuracy than the addition of GR4J SM alone (although the combination of all three soil moisture data sources resulted in the greatest improvement overall). It is also evident in Table 17, that for all catchments, the inclusion of ECV SM and POAMA SM resulted in improved reliability and resolution of the forecast distributions. As such, it is considered that the best overall results in terms of median accuracy and forecast reliability and precision was achieved with the combination of all soil moisture inputs.

Table 17 Comparison of the different models considered for the validation period on each catchment

	NSE	RMSE	Bias (%)	α	$\pi^{(rel)}$	Hit Rate (%)
			A2390512			
base	0.78	4376	12%	0.75	1.42	97%
base+GR4J SM	0.89	3075	10%	0.73	1.48	100%
base+ECV SM+POAMA SM	0.74	4766	15%	0.79	1.51	93%
base+all SM	0.88	3262	10%	0.75	1.45	100%
GR4J with POAMA rainfall	-0.16	10145	-38%	0.85	1.32	73%
			A2390514			
base	0.68	2663	-29%	0.77	0.44	100%
base+GR4J SM	0.75	2340	-22%	0.74	0.49	100%
base+ECV SM+POAMA SM	0.68	2669	-37%	0.82	0.44	100%
base+all SM	0.67	2729	-21%	0.82	0.60	100%
GR4J with POAMA rainfall	-0.38	5642	28%	0.75	0.91	87%
			A2390519			
base	0.00	2772	86%	0.42	0.79	100%
base+GR4J SM	0.20	2481	86%	0.41	0.83	100%
base+ECV SM+POAMA SM	0.30	2320	64%	0.46	0.81	100%
base+all SM	0.46	2043	64%	0.44	0.84	100%
GR4J with POAMA rainfall	-2.08	4987	53%	0.72	0.86	97%

When comparing the ANN models (both with and without the GR4J input data) to the GR4J model, it can be seen that the GR4J model performed relatively poorly on all catchments in terms of median forecast accuracy. Furthermore, while the reliability and resolution of the forecast distributions obtained using the GR4J model seem fairly good when compared with the ANN results, hit rates of only 73% and 87% were obtained for catchments A2390512 and A2390514, respectively, indicating that the forecast 95th percentile bounds were underestimated using this model. For catchment A2390512, the forecast predictive distribution obtained using the GR4J model were the widest of all of the models considered (as evidenced by the smallest $\pi^{(rel)}$ value), yet this model also resulted in the lowest (worst) hit rate. The relative performances of the ANN and GR4J models can be compared visually in Figures Figure 48 and Figure 49. In Figure 48, the median forecasts and 95% forecast limits were generated using the ***base+all SM*** models for each catchment as these models were considered to give the best forecasts overall. Figure 49 shows the same outputs generated by the GR4J model in comparison to the observed flows. It can be seen that the forecasts produced by the ANN models give a much better representation of the observed monthly flows in comparison to those produced by the GR4J model. Furthermore, the GR4J model forecasts tend to have greater associated uncertainty as evidenced by the widths of the 95% uncertainty bounds.

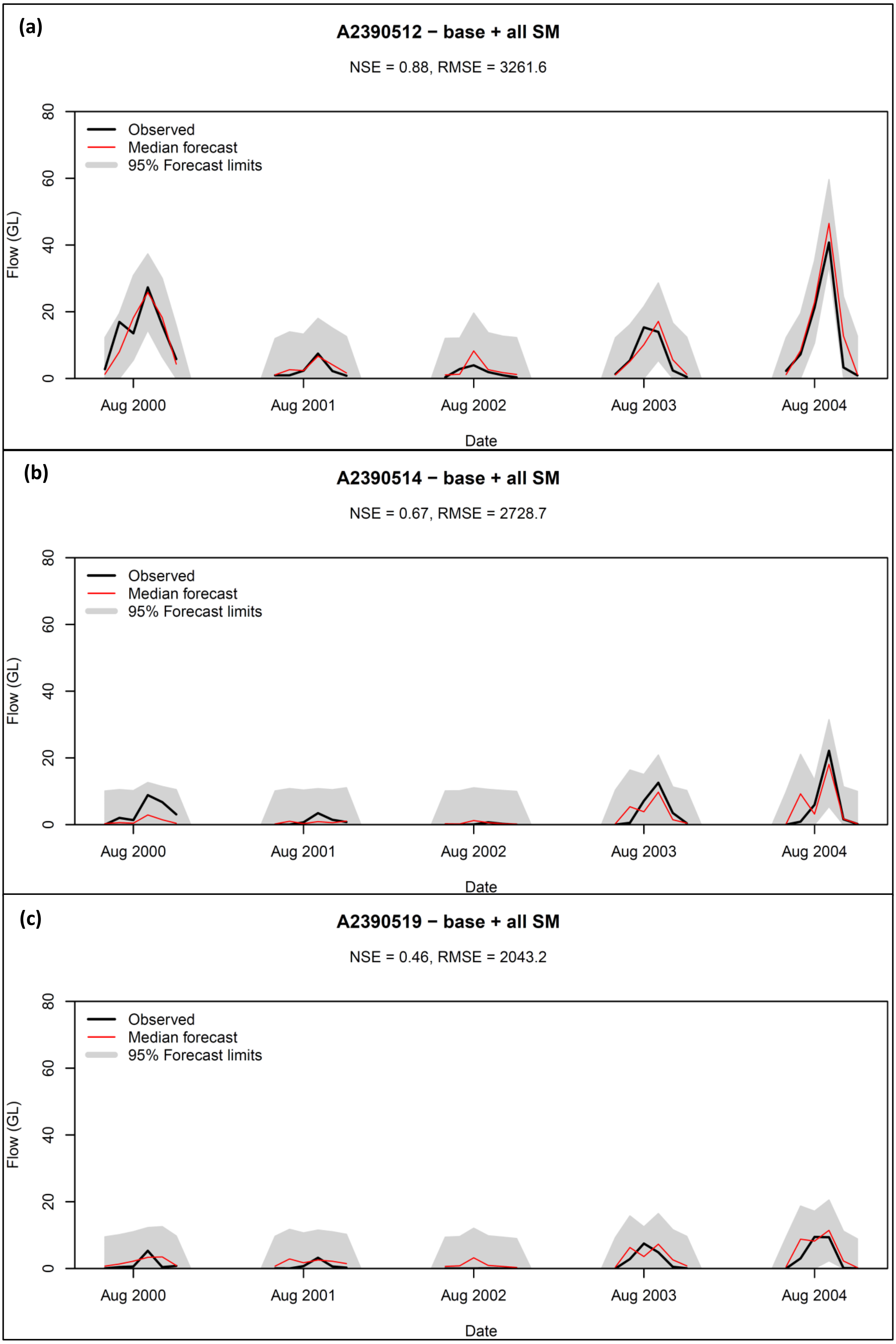


Figure 48 Median monthly flow forecasts (red) and 95% uncertainty bounds (grey shaded) generated using the *base+all SM* ANN models in comparison to observed flows (black) at gauges (a) A2390512, (b) A2390514 and (c) A2390519.

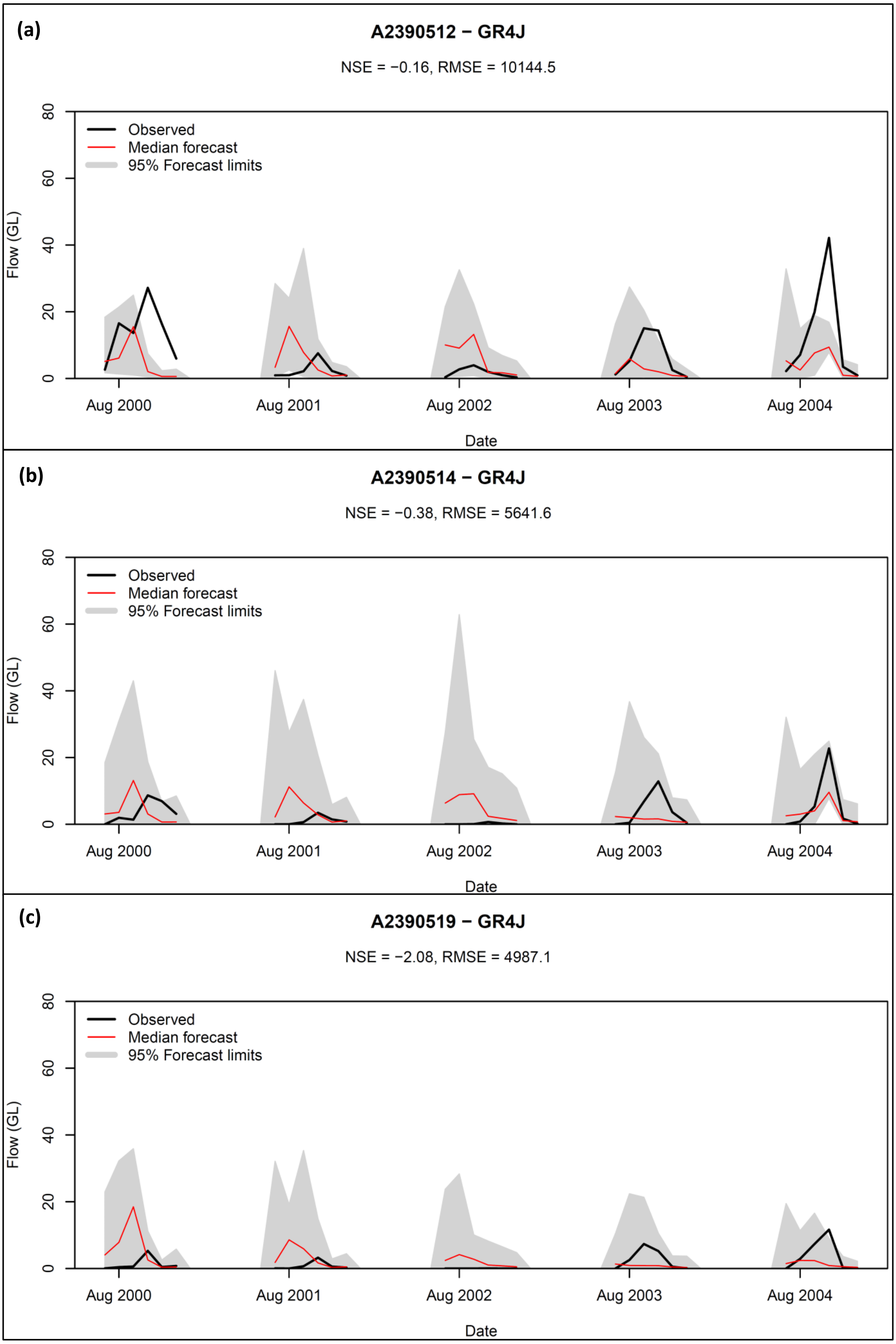


Figure 49 Median monthly flow forecasts (red) and 95% uncertainty bounds (grey shaded) generated using the GR4J model in comparison to observed flows (black) at gauges (a) A2390512, (b) A2390514 and (c) A2390519.

Given the results presented in Section 6.1, showing the reduction in median model accuracy when an uncertain climate forecast (POAMA) is used to inform the GR4J model in comparison to when a perfect climate forecast is available (Actual), it is not surprising that the ANN models are more accurate than the GR4J models. This is because the GR4J model relies solely on the rainfall and evapotranspiration inputs (together with the model structure, initial states and calibrated parameters) to generate the forecasts. For the ANN models, on the other hand, the uncertain and potentially erroneous climate forecast is only one of at least five model inputs (see Table 13 - 15 and Table 9 for the number of selected inputs in each model), where the remaining inputs include a number of observed variables.

The ANN forecasts are based on the same input data as those used to calibrate the models, i.e. the POAMA climate data (albeit from different time periods). In comparison, the GR4J models were calibrated to actual daily rainfall and evapotranspiration data from the calibration period. As such, the ANN models are calibrated to correct for any bias or consistent error in the POAMA forecast or downscaling method, where the CRRs assume that the analogue period identified by downscaling the POAMA forecasts represents the future rainfall. As such, if the forecasted climate data does not accurately represent the observed climate data, the calibrated model parameters may not be appropriate. There may also be a temporal aspect that effects the CRRs, whereas these models simulate flow at a daily time step, even if the monthly total rainfall is similar, the distribution of this rainfall over the month will influence this simulated streamflow. The GR4J SM data used in the ANN models are not similarly affected, as these are based on the calibration results using observed rainfall and evapotranspiration data.

Overall, all models (ANN and GR4J models), performed relatively poorly when applied to forecast flows at gauge A2390519. As discussed in Section 6.1, in the case of the GR4J model, this was likely due to the unrepresentative data used for calibration (i.e. the model was calibrated to a wetter runoff regime than the drier regime experienced during the validation period). For the ANN models, this result is likely due to the lack of data (e.g. groundwater data) needed for appropriately characterising the monthly flows. Nevertheless, the ***base+all SM*** ANN model is more informative about the expected flow in the upcoming month than the historical monthly flow distribution, as can be seen in Figure 50 for both the calibration and validation periods.

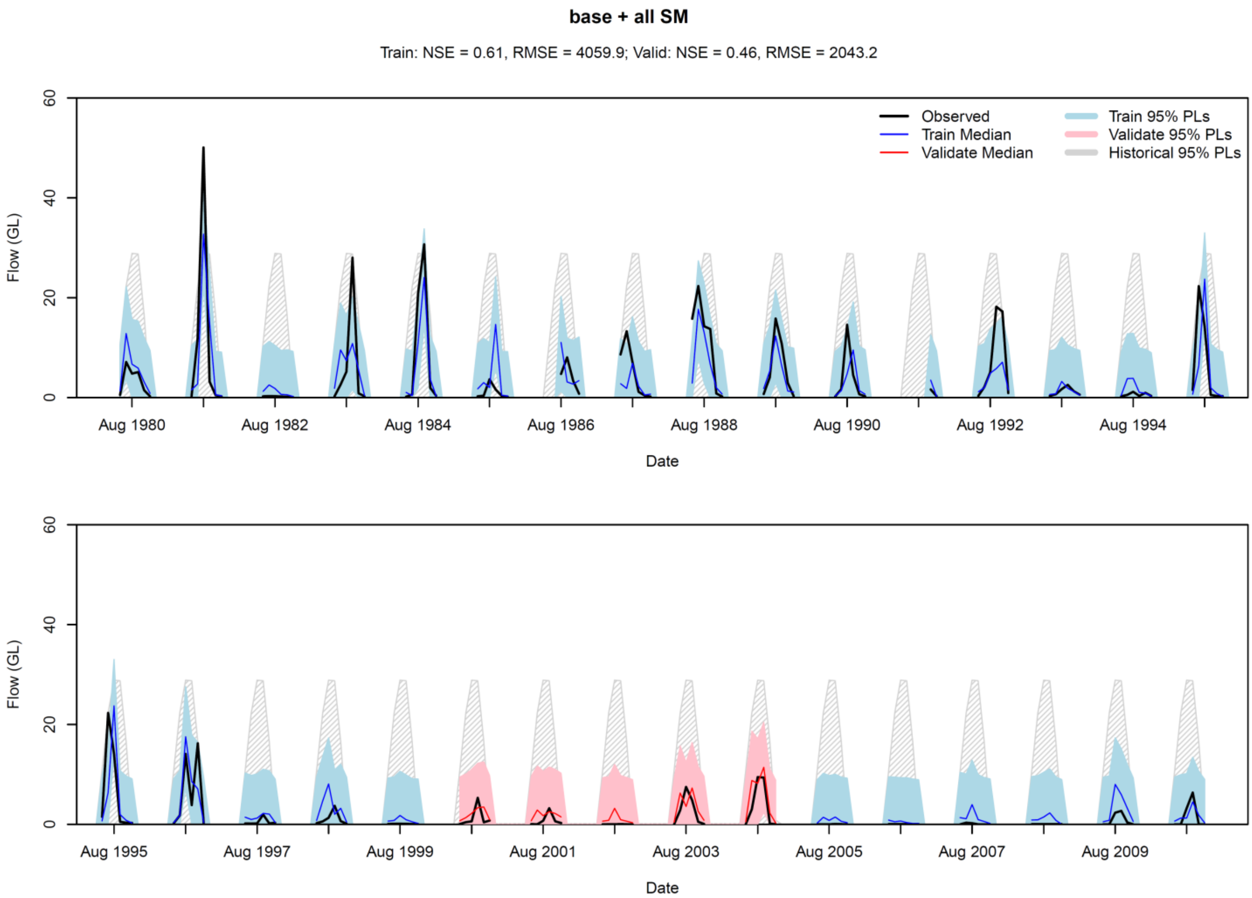


Figure 50 A2390519 median monthly flow forecasts and 95% uncertainty bounds (blue and red for calibration and validation periods, respectively) generated using the *base+all SM* ANN model in comparison to observed flows (black) and historical monthly flow distributions.

While the sample size of three catchments is still relatively small, it can be seen that the ANN model with appropriate inputs (i.e. including all soil moisture inputs) produced relatively robust results across the three catchments. This can be contrasted to the GR4J model alone, which provided significantly less accurate and more uncertain forecasts, largely due to the large dependence on the uncertain and potentially erroneous POAMA forecasts, particularly at the daily time step.

Therefore, should all required input data be available in real time for forecasting purposes, the results suggest that an ANN model with all soil moisture inputs is most suited to the providing flow forecasts for the Drain M catchments. However, the availability of the additional input data required by the ANN models (in comparison to the GR4J model which requires only rainfall and evapotranspiration inputs) depends on telemetry for groundwater wells, and the availability of the derived ECV soil moisture product from raw real time remotely sensed data.

7 2014 Implementation Trial

The 2014 flow season has been used to test the models developed in forecasting streamflow in Drain M. Despite the results from Section 6, the CRR model alone was used in the 2014 trial. This was due to a number of reasons, including the ECV soil moisture dataset was not available in timeframes necessary to provide monthly streamflow forecasts, the 2014 trial application was undertaken in parallel with model development, and the CRR models alone provided comparable results to the ANN models in terms of the reliability of the predictive distribution. Remotely sensed soil moisture data is likely to become more readily available in the future, for example NASA's Soil Moisture Active Passive (SMAP) instrument was launched in early 2015, and has both active and passive sensors to produce the highest-resolution, most accurate measurements ever made of soil moisture (http://smap.jpl.nasa.gov/).

The steps involved in producing a streamflow forecast using the CRRs are as follows:

1. Update and process model input and output datasets from SILO (rainfall and PET) and Hydstra (streamflow)
2. Estimate the GR4J parameter distributions based on the past 10 years of data (at the start of the season only)
3. Estimate residual error model parameters for each sampled set of GR4J model parameters from the distributions identified
4. Download POAMA climate forecast
5. Downscale POAMA forecasts using the analogue approach
6. Run climate forecasts through GR4J models
7. Plot outputs.

R scripts to automate each step in the process have been developed and archived in the DEWNR model warehouse. Readme files to run the scripts were provided as part of the package. This section provides a summary of the results from the 2014 trial.

Unfortunately, the streamflow recorded in Drain M in 2014 was limited, with one short flow event in August recorded at gauges A2390514 and A2390519, and flow over July and August at the most downstream gauge of A2390512. This came after close to median rainfall for the first six months of the year, and then near lowest rainfall on record in August and September, resulting in no more flow in Drain M for the year (Figure 51). The observed flow is discussed further in Section 7.2.

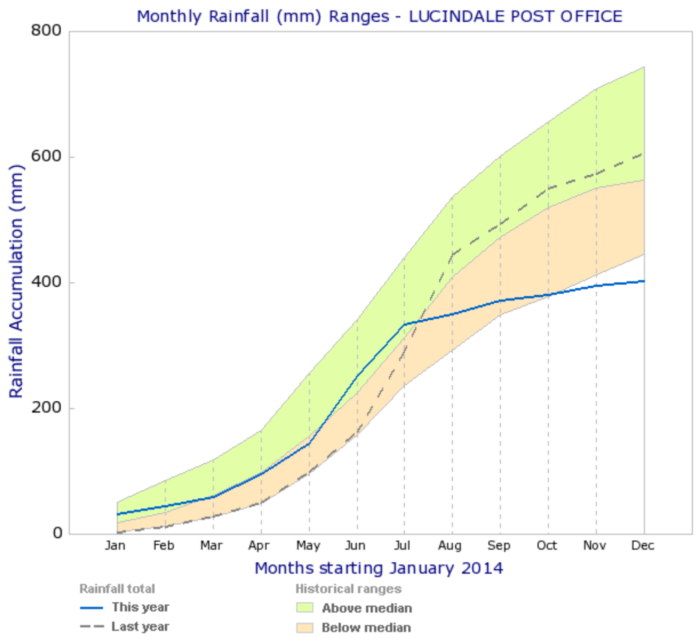


Figure 51 Cumulative monthly rainfall at Lucindale rainfall station. 2014 rainfall can be seen to track close to median (between green and red areas) until July, and then flatten off substantially, indicating minimal rainfall over July – December (Bureau of Meteorology, http://www.bom.gov.au/watl/rainfall/ranges.shtml).

7.1 Rainfall Forecasts

The POAMA rainfall forecasts, after applying the analogue downscaling approach of Shao and Li (2013) (Section 0), are presented in Figure 52. There are 33 POAMA model outputs available for each forecast period (11 runs each from 3 different models), with the monthly rainfall forecast from these outputs presented as the boxplots in Figure 52. The accuracy of the forecast can be compared to the observed rainfall that occurred for each month in each catchment (calculated as the average across the catchment using a Theissen polygon approach) as the red dot, and for comparison the long term average rainfall is presented as the green dot.

It can be seen from Figure 52 that near average rainfall was received for each month in 2014 from January to July, as the red dot is close to the green dot. In comparison, the remainder of the year was significantly below average. The POAMA rainfall model did not forecast these very dry months, with the distribution of the forecasts (represented by the boxplot) centred on the historical average. This can be seen in more detail for the September forecast in Figure 53, with the cumulative distribution in rainfall for the month forecast (green line) very similar to the distribution historically for the month (red line). This can be considered an uninformed forecast, where forecast distribution of rainfall in this September is the same as the distribution of rainfall that has occurred historically in September. It can be seen from the red lines in Figure 53 that the probability of getting less than the approximately 25 mm of rainfall recorded in September 2014 is very low (either forecast or historically).

To ensure that this result is not an artefact of the downscaling approach that has been applied in this work, the results presented in Figure 52 and Figure 53 have been compared to the seasonal rainfall forecast provided by the Bureau of Meteorology (Figure 54). The results are not directly comparable, as Figure 53 presents the forecast for the month of September, where Figure 54 presents the forecast for the three month period September to November. It can be seen from Figure 54 that for the South East of South Australia the probability of median rainfall was forecast to be close to 50%, which is similar to that in Figure 53, where the two lines are close at the 50th percentile rainfall.

To check the remainder of the three month season considered in Figure 54 (where the seasonal rainfall could have returned to close to median after a dry start), the three month rainfall totals for the three catchments have been considered. In 2014, the September to November rainfall was approximately 65 mm for the catchment area represented by the A2390514 gauge. For this catchment area, this was the lowest total over the period used to calculate the median in Figure 54 (1981-2010), and the second lowest over the 126 year rainfall period in the SILO rainfall dataset. The result was similar for the other two areas considered, with the period being the fourth lowest in the SILO record for the A2390512 catchment area, and the fifth lowest on record for the A2390519 catchment. Given this, it is likely that this very dry season was not forecast by the POAMA rainfall model, rather than an inconsistency with the downscaling approach used.

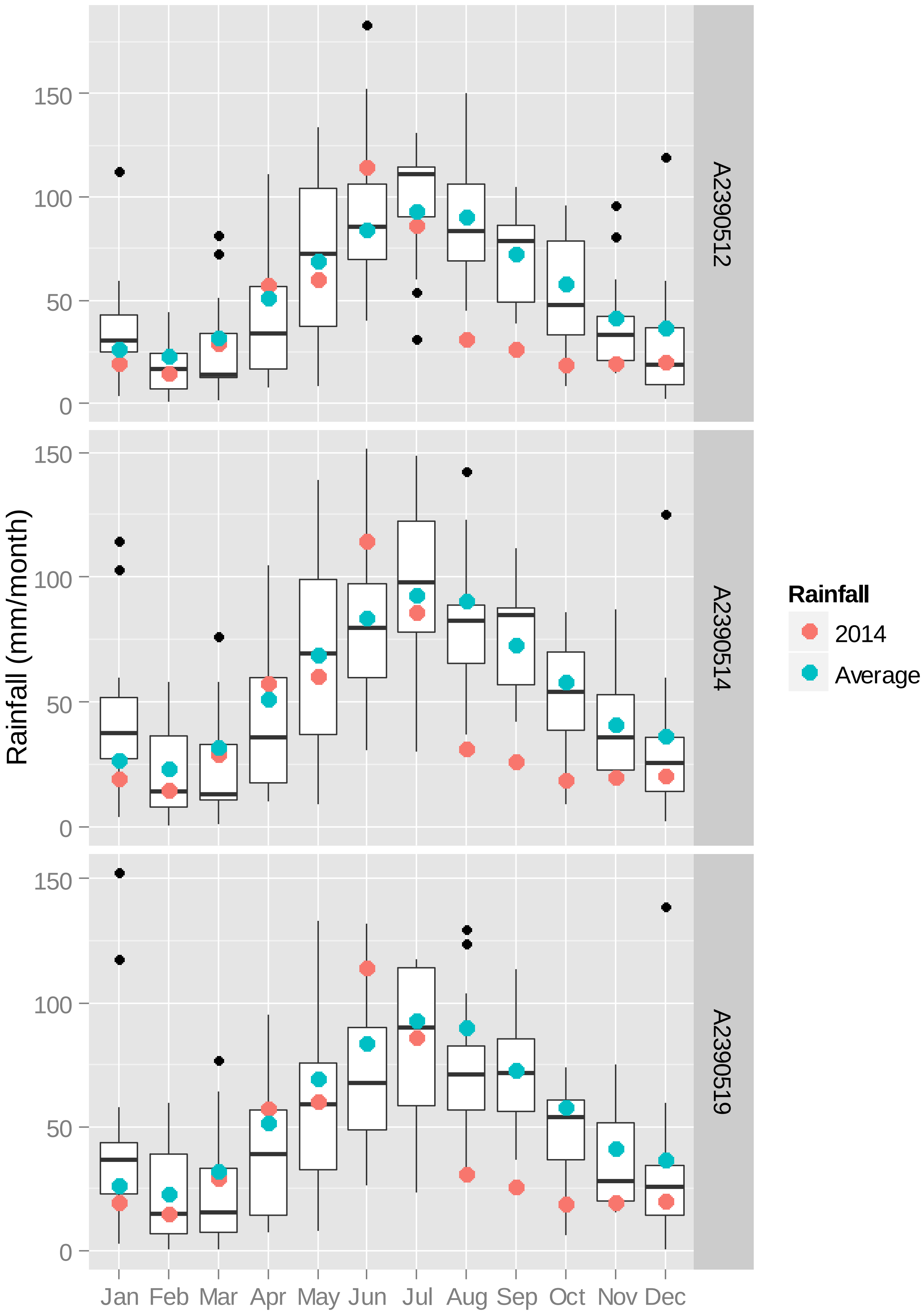


Figure 52 Boxplots represent the downscaled POAMA forecast each month for 2014, with the observed monthly total in red and the long term average in green. It can be seen that the downscaled POAMA forecast substantially overestimated rainfall in August and September 2014.

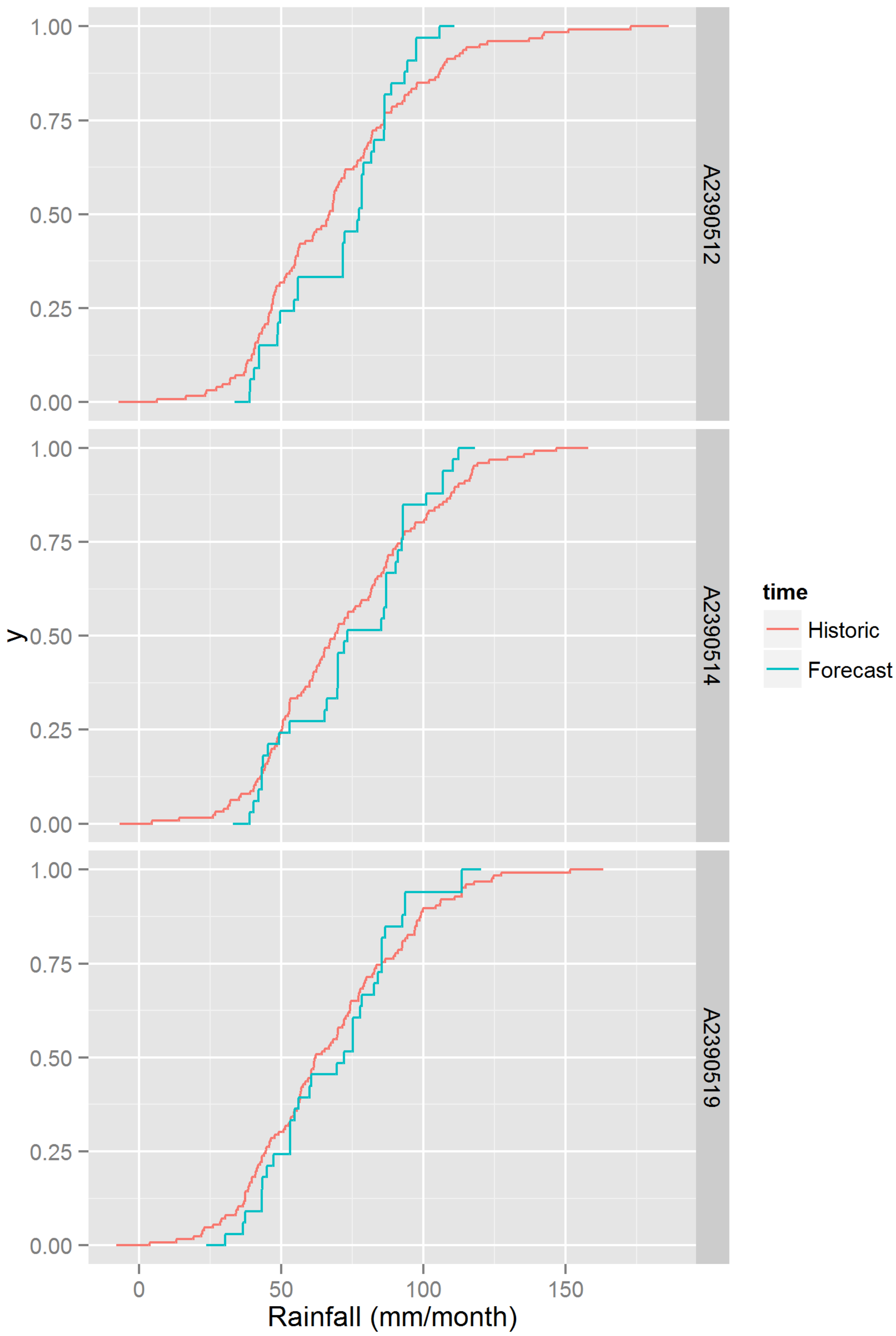


Figure 53 Rainfall Forecast for September, compared to the historical distribution of September monthly rainfall. The y-axis can be read as "the probability that the monthly rainfall is less than y".

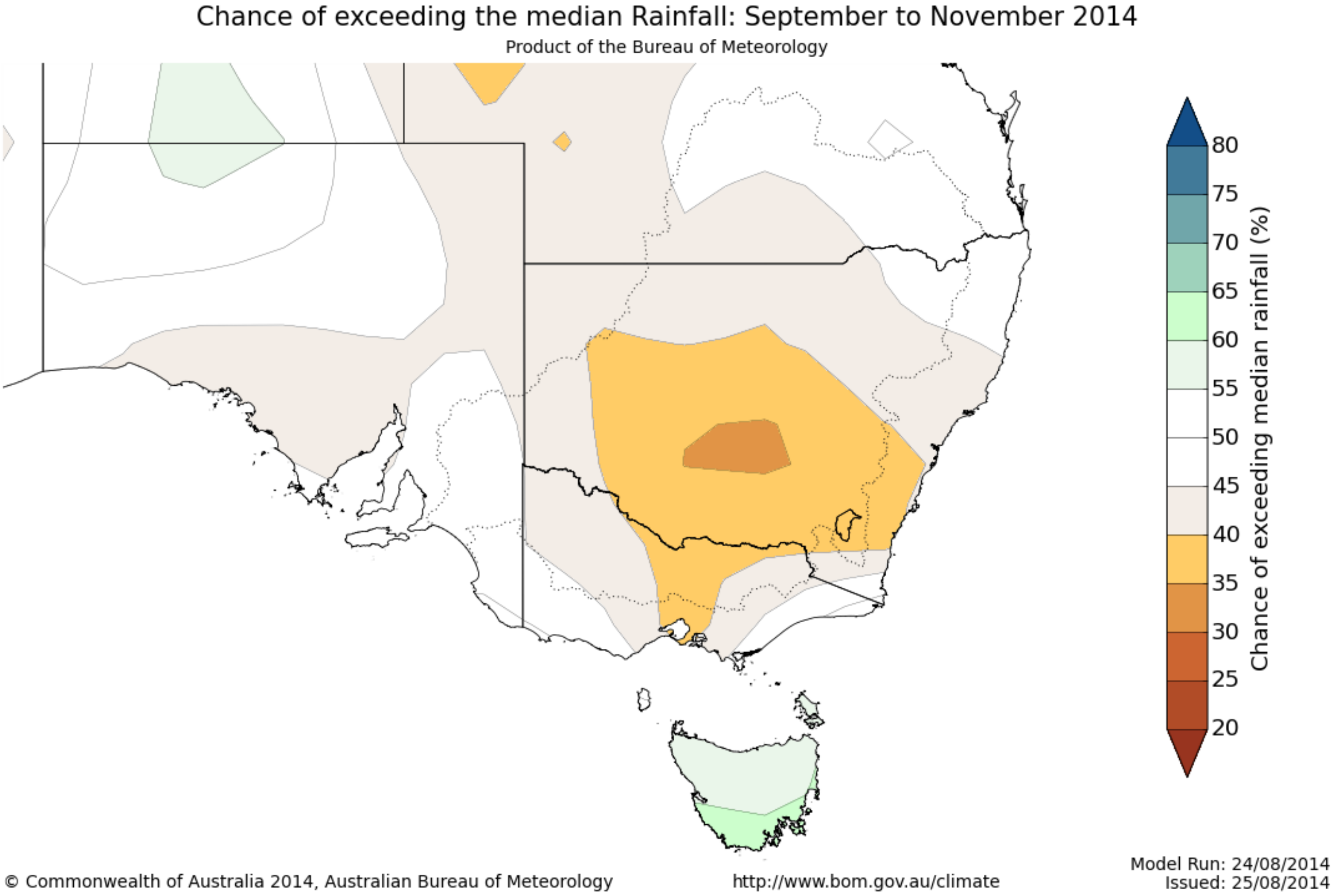


Figure 54 Seasonal Rainfall forecast for September to November 2014, with close to a 50% chance of exceeding median rainfall for the South East of South Australia (Bureau of Meteorology, http://www.bom.gov.au/climate/outlooks/#/rainfall/median/seasonal/0).

7.2 Streamflow Model and Forecast

Given that the rainfall forecast substantially over predicted the observed rainfall in the second half of 2014, the streamflow model predictions also substantially overestimated the observed flow. An example of the outputs produced for September is provided in Figure 55 as a cumulative distribution. The antecedent conditions after a very dry August were also dry, which are captured by using the previous recorded rainfall to initialise the production store in GR4J, and using the observed flow on the day of the forecast to initialise the routing store in the model. Therefore, even though the rainfall forecast was for a distribution similar to the historical distribution (Figure 53), the streamflow forecast distribution was much lower than the historical distribution, but still overestimated the observed streamflow of 72 ML/d at gauge A2390512, and no flow at the two other gauges.

To compare the runoff models without the influence of the POAMA rainfall forecast, the models have been run with the observed rainfall and PET as the input (Figure 56). It can be seen that the models on all catchments tended to start to generate flow June, before any runoff was observed in the catchment. This may be an issue with the GR4J model, where a more complex model structure may be able to represent the break of season more accurately. This component of the flow regime may also be able to be represented better by adopting a different likelihood function used to infer the model parameters. However, the function adopted does attempt to balance the effect of the range in the flow regime on the likelihood calculated, as the daily flows have a square root transform applied, and the percentiles of the Flow Duration Curve are also considered.

The models did forecast the flow when the A2390512 catchment did start to flow in July, albeit the observed flow was at the lower end of the forecasts. For the two other catchments, the model tended to simulate some flow in response to rainfall, where for both catchments there was only the one flow event in August. This event was relatively well simulated for A2390519, but overestimated for A2390514. Based on the results presented in Section

6, it is likely that the ANN model with the GR4J storage as an input would provide a more accurate representation of the monthly streamflow for these two catchments, in particular the dry period of A2390514 catchment (given that the GR4J model had a very high bias in Table 17).

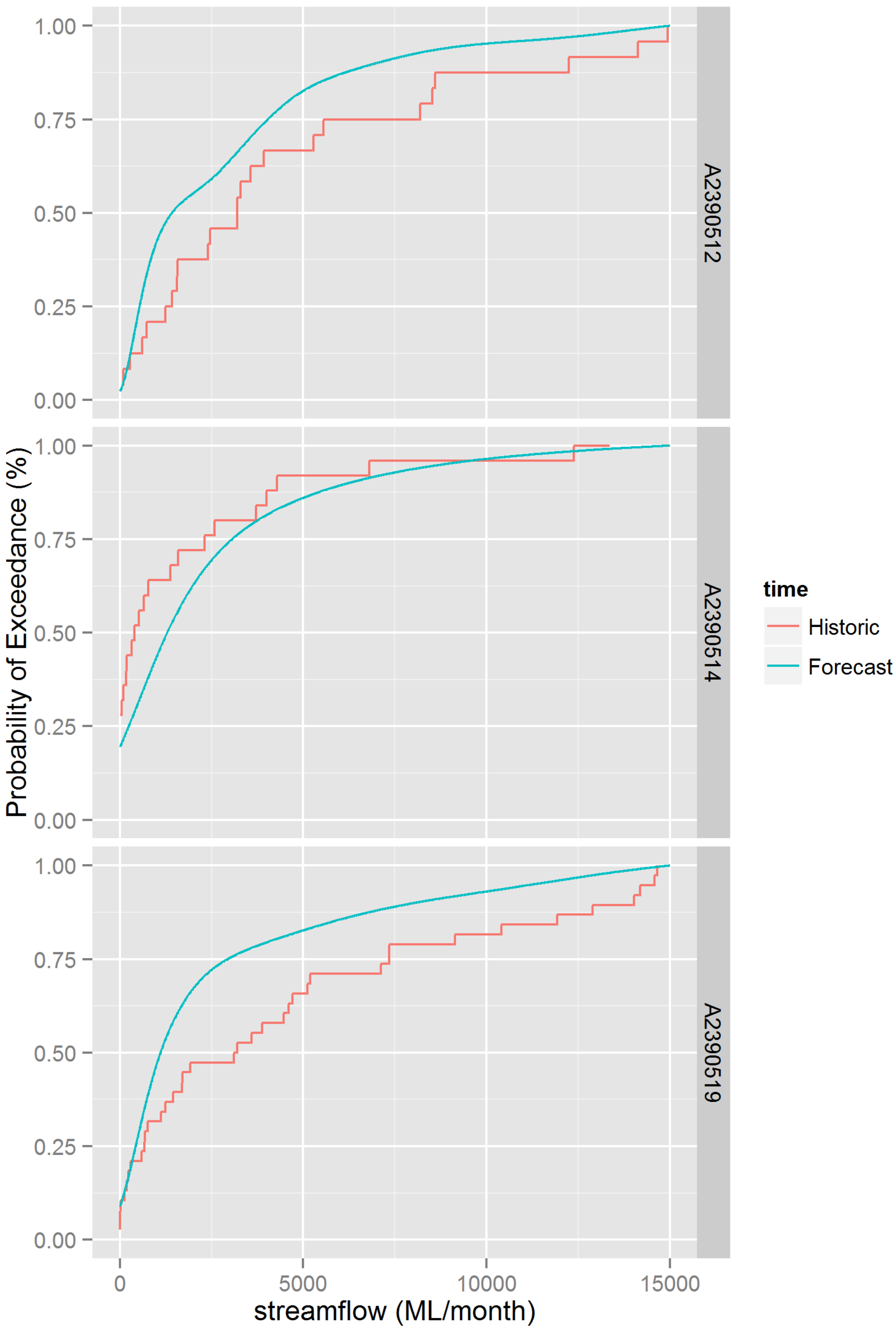


Figure 55 Even though the reduced rainfall was not forecast, the models did forecast less runoff was likely compared to the historical distribution of rainfall for September, due to the antecedent catchment conditions

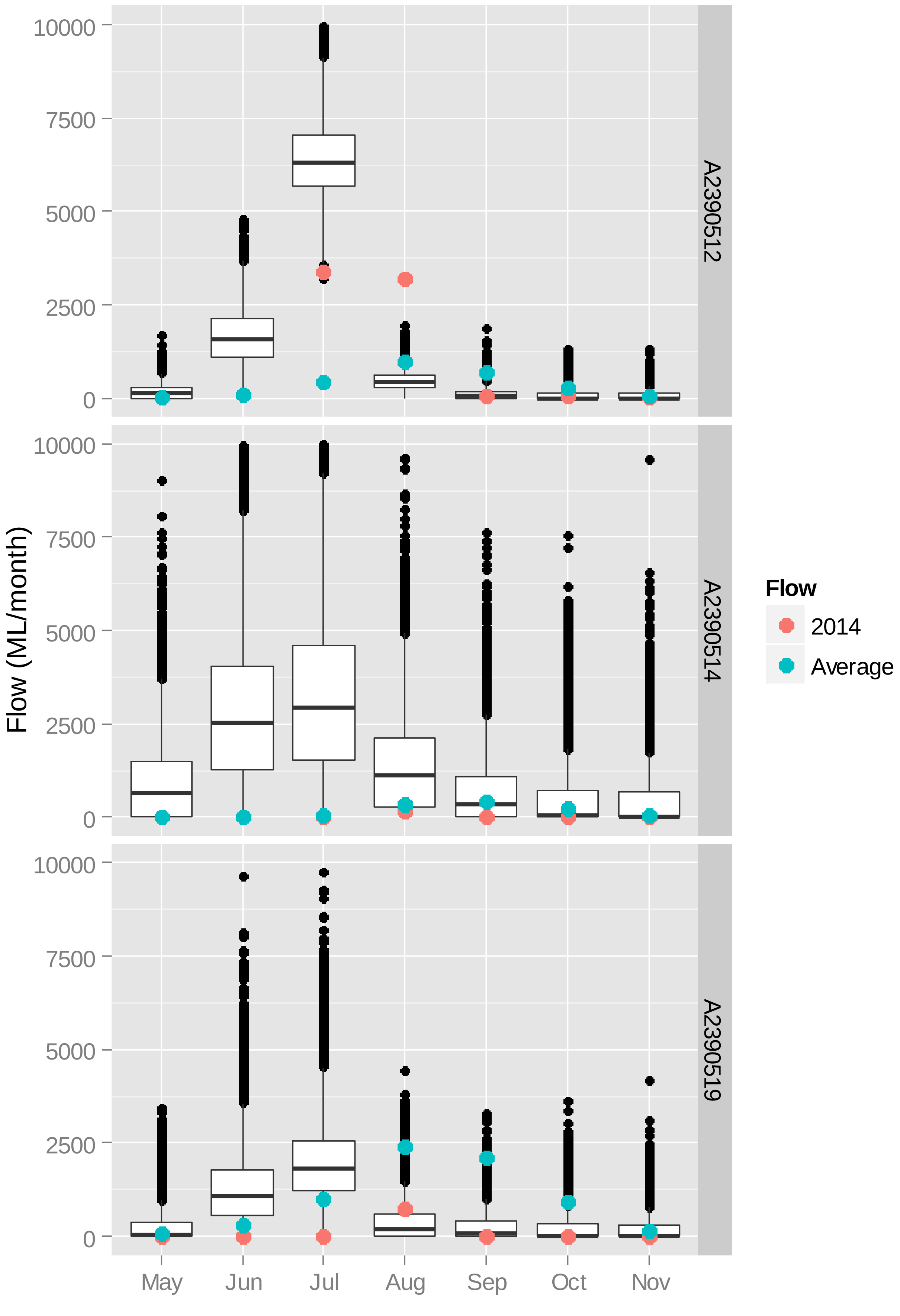


Figure 56 57 Boxplots represent the downscaled POAMA forecast each month for 2014, with the observed monthly total in red and the long term average in green. It can be seen that the downscaled POAMA forecast substantially overestimated rainfall in August and September 2014.

7.3 Summary and Recommendations

- The value of information on antecedent conditions is clear from this work, where the assimilation and warmup of the GR4J model improve the reliability of the forecasts compared to the rainfall forecast for the extremely dry months experienced.
- While not perfect, the models did estimate the flow events that did occur based on observed rainfall well, indicating that the models developed are likely to have some use in informing the management of the drainage network.
- However, even when observed rainfall was used as input to the runoff model, the flow at the start of the season was overestimated. A different model structure could be considered to attempt to have a better representation of the commence-to-flow for the season, and low flows more generally. From a management perspective, accurate prediction of the low flow months may be of less importance, and it is advance notice of large volumes that are likely to occur which are of most interest for the management of the system and the potential to divert flows around the system.
- In the previous section, the statistical models were found to provide much better performance than the CRRs, particularly in metrics that represent the accuracy of the median of the model forecast. One of the main inputs to the statistical model was identified to be remotely sensed soil moisture data. Currently, post-processed remotely sensed soil moisture products, which aggregate both passive and actively measured information, are not available in real time for streamflow forecasting. However, this information is likely to become more readily available in the future, for example NASA's Soil Moisture Active Passive (SMAP) instrument was launched in early 2015, and has both active and passive sensors to produce the highest-resolution, most accurate measurements ever made of soil moisture (http://smap.jpl.nasa.gov/). As this data becomes more readily available in near real time, it is likely to be useful to make streamflow forecasts more accurate, either through statistical models, or as assimilation data for lumped conceptual models (or both).
- Over the long term, the downscaled POAMA rainfall forecasts were found to be reliable. However, the monthly rainfall was substantially overestimated in some months during the 2014 trial. While incorporating the influence of antecedent conditions in the catchment has been demonstrated to improve the results, the accuracy of streamflow forecasts are directly linked to the accuracy of rainfall forecasts. Seasonal rainfall forecasts for Australia, available for research purposes from the Bureau of Meteorology, haven been used, and an approach to correct for local rainfall influences based on the larger scale climate model (downscaling) published by the Bureau of Meteorology have been adopted in this trial. These climate models will continue to improve, and provide more reliable rainfall forecasts. Other approaches to downscale the model outputs, or correct for any forecast errors (such as using the forecast data in the model calibration, as was the case for the ANNs), may also improve the rainfall forecast, and in turn the streamflow forecast.

8 Water Balance Modelling

The modelled streamflow forecasts developed provide a prediction of the flow likely to be available in an upcoming month. To assess potential benefits and impacts of different diversion scenarios on the wetlands in the Drain M system, a water balance model has been developed. The water balance model allows different operational scenarios to be considered, and resulting water level information to be estimated based on the forecasts provided. The water level results are of particular interest as this parameter is the basis of environmental water requirements and rules around releases from Bool Lagoon.

8.1 Network schematic

An eWater Source model has been developed as the water balance model, with the schematic of the model for Drain M seen in Figure 58. Two storages (blue triangle) are represented in the model, Bool Lagoon and Lake George. Details on the data and calibration of these two storages are provided in the next section.

Diversion and release rules are input to the model using the yellow minimum flow requirement nodes. These nodes are used to represent the management rules for releases from Bool Lagoon and Lake George based on water levels, and rules to divert water from Drain M along the REFLOWS channel. Input to these nodes allow different management options to be considered.

The required flow inputs are:

- Mosquito Creek - inflows to Bool Lagoon (A2390519). It should be noted that other small inflows to Bool lagoon (e.g. Seymour Robertson drain and other private drains) are not explicitly accounted for
- Bakers Range South - all drains between Bool Lagoon and Callendale (A2390514)
- DSCallendale – Catchment area downstream of Callendale (A2390512)
- Sutherlands Drain – calculated in the model as 1.105% of the flow at node A2390512 (based on the contributing areas)
- Groundwater - the groundwater input to Lake George based on the results of Australian Water Environments (2009) (See section 8.3.4)
- OceanInflow - flow required to maintain water levels when gates at the outlet of Lake George were open (not a necessary input in forecast mode)

As such, three flow inputs are required to run the model, the flow forecasts at locations A2390519, A2390514 and A2390512.

The climate inputs to the forecast models are also required as inputs for the reaches and storages, to calculate the net evapotranspiration. Travel times are calibrated as part of the forecast runoff model, and as such are not recalculated in the routing model.

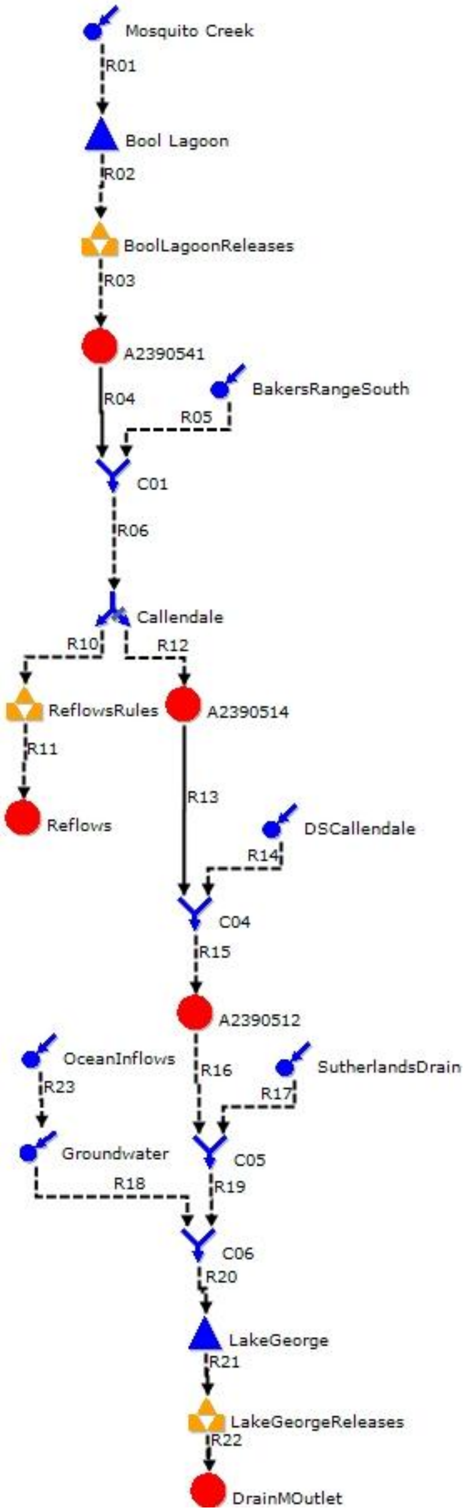


Figure 58 Network schematic

There are two more elements in the model that include input data and calibrated values:

- Bool and Hacks Lagoon
- Lake George

8.2 Bool and Hacks Lagoons

8.2.1 Dimensions

The 2m DEM with any discontinuities created by flight paths repaired (undertaken by AAMHatch in 2012) was used to develop the depth – area – volume relationship for the model. The DEM was clipped by the South Australian Wetland Inventory Directory (SAWID) polygon boundary for both Bool and Hacks Lagoon, before the volume and area was calculated for depths in steps of 1 cm using the 3D analyst tool in ArcGIS, automated by a Python script.

The shape of the relationship between depth and area or depth and volume derived from the DEM are expected to be accurate. However, the absolute value for the depth (and corresponding area and volume) is important for the model, as the release rules are based on the water surface elevation in different months of the year. Elevations from the DEM were compared to a total of 2345 survey points available within Bool and Hacks Lagoon, collected in 2011. It was assumed that the on ground survey points are more accurate than the LiDAR derived DEM. The exceedance curve of the difference between the DEM and the survey points can be seen in Figure 59, where the DEM can be seen to be consistently higher than the survey points, where only 25% of the DEM pixels were below the survey point located within the pixel, compared to 75% above. Given this positive bias, the DEM was shifted down 5 cm so the median of differences between the DEM and the survey points across the 2345 points was zero. This shift results in 78% of the comparison points within 10 cm of the surveyed height, and 88% of the comparison points within 15 cm of the surveyed height. Based on this shifted DEM, the relationship between level and area, and level and volume, is presented in Figure 60.

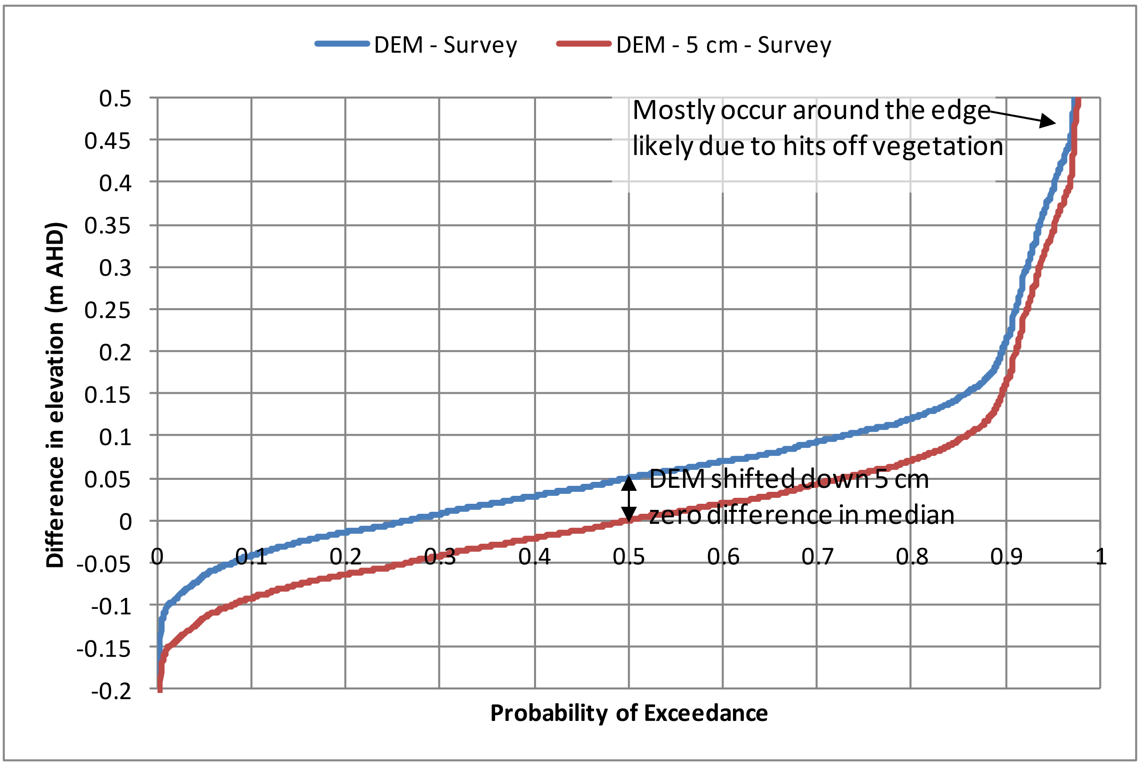


Figure 59 Comparison between surveyed elevations and 2m DEM

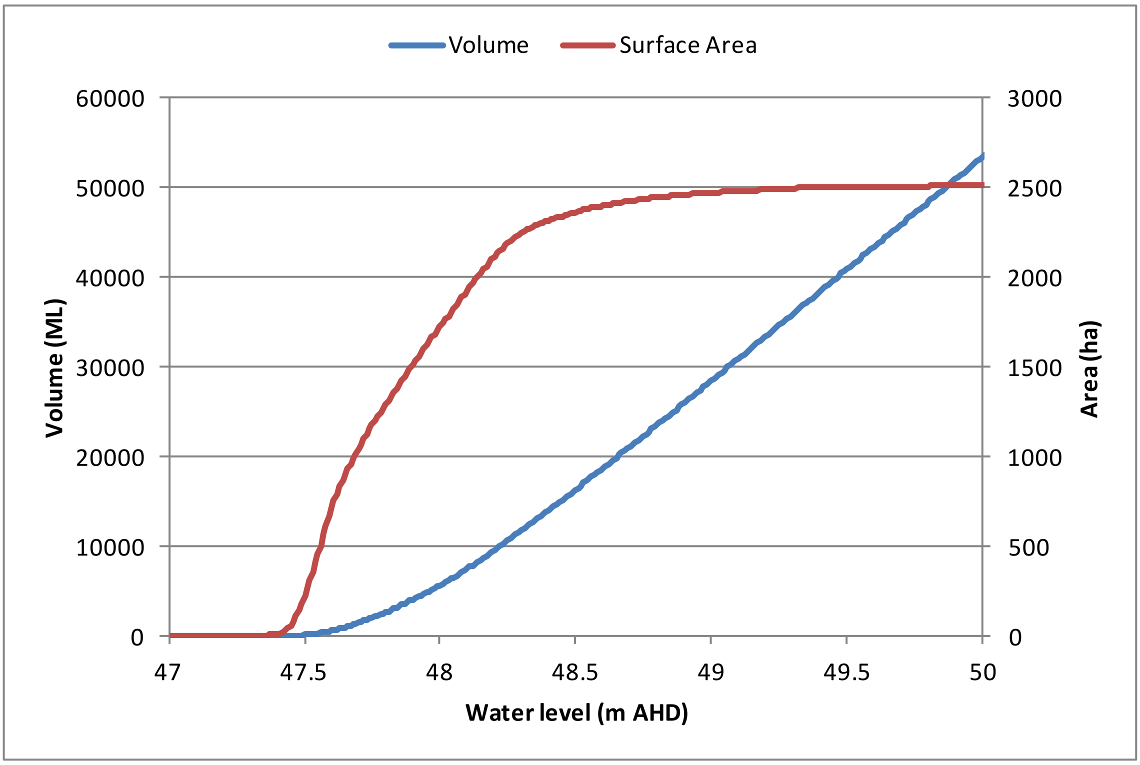


Figure 60 Bool Lagoon hypsometric curves

8.2.2 Climate Data

SILO data at Struan (26082) has been used for the rainfall and potential evapotranspiration (PET) data for Bool Lagoon. The rainfall data provided for the site has been adopted given its proximity to Bool Lagoon, however the site has been closed, and hence data interpolated from other sites, since 1999. Class A evaporation has been used as the evaporation data with a pan factor of 0.8, based on previous studies at Lake George (Australian Water Environments, 2009) and the Coorong (Webster, 2005).

8.2.3 Release Rules

The release rules outlined in the Bool Lagoon Management Plan (Department for Environment and Heritage, 2006a) have been implemented on the Minimum Flow requirement node downstream of Bool Lagoon (Figure 58). The release volume each day is calculated to maintain the water level at the following:

- 48.15 m AHD during June;
- 48.24 m AHD during July;
- 48.30 m AHD at the end of the first week in August;
- 48.40 m AHD during the third week of August;
- 48.55 m AHD at the end of August; and
- 48.61 m AHD for as long as possible from the second week in September.

No release occurs if the water level in the wetland is below the maximum level for the relevant time of year. A maximum release rate of 1000 ML/d was also implemented. While there are days with flow greater than this in the observed data, the maximum release rate was calibrated based on the daily NSE between the modelled flow and that recorded at the Bool Lagoon outlet, at station A2390541.

8.2.4 Seepage Rate

The seepage rate was also calibrated based on a comparison between the modelled and gauged flow downstream of Bool Lagoon. A constant seepage rate of 10 mm/d has been used, to provide a small difference in total volume released between the modelled and recorded data. This losing condition might be expected, based on groundwater well ROB005 located on the edge of Bool Lagoon, where the depth to groundwater is generally greater than 1 m (Figure 61).

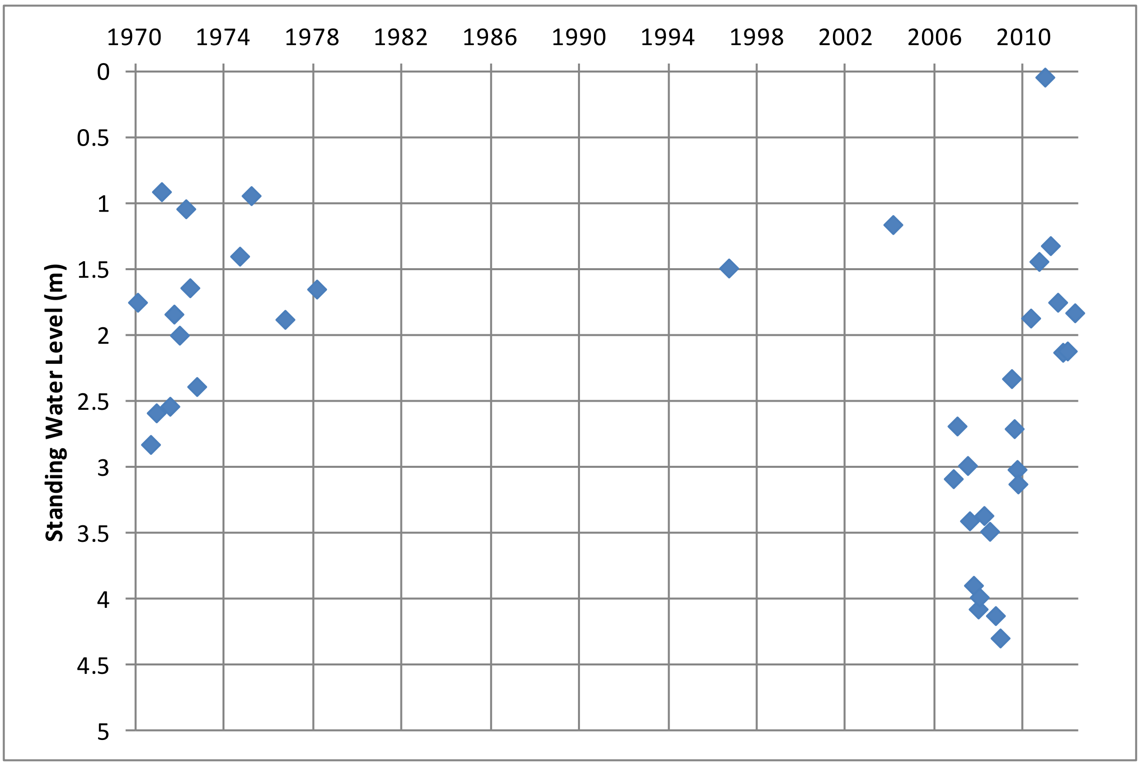


Figure 61 Standing water level for well ROB005, located on the edge of Bool Lagoon.

8.2.5 Calibration Results

The daily flow downstream of Bool Lagoon, both modelled and recorded at gauge A2390541 is presented in Figure 62. The periods where releases occur, and the volume that was released, is considered a surrogate of the accuracy of the Bool Lagoon storage model, as the water level must exceed (or be expected to exceed) that in the management plan for releases to occur, and the comparison of volumes also indicates if a representative storage relationship has been adopted, including losses (evaporation and seepage) and gains (inflow and rainfall). The inflows to the storage are considered accurate, given the gauge on Mosquito Creek at Struan (A2390519) is directly upstream. The NSE of the daily flows is 0.54 (used to calibrate the maximum release rate) and the total volume bias is 1.7% (used to calibrate the seepage rate). However, daily flows are also driven by management decisions, and much of the inaccuracy in the NSE value is caused by the model stopping releases abruptly, where the gauged data suggests flows are managed to recede more slowly. As such, it is the ability of the model to represent the occurrence (and non-occurrence of release events that is of most interest, along with the volume that is released over the event, rather than attempting to replicate the daily flows recorded downstream of Bool lagoon. A comparison of annual volumes is presented in Figure 63, where a R^2 value of 0.78 was obtained at the annual time scale.

It can be seen in Figure 62 that the recent periods of multiple years without flows has resulted in modelled releases when they have not actually occurred. This is likely to be a combination of dry initial conditions developing over a number of years not being represented in the model, as well as lower than usual groundwater levels, for example more than 4m deep in 2009 (Figure 61).

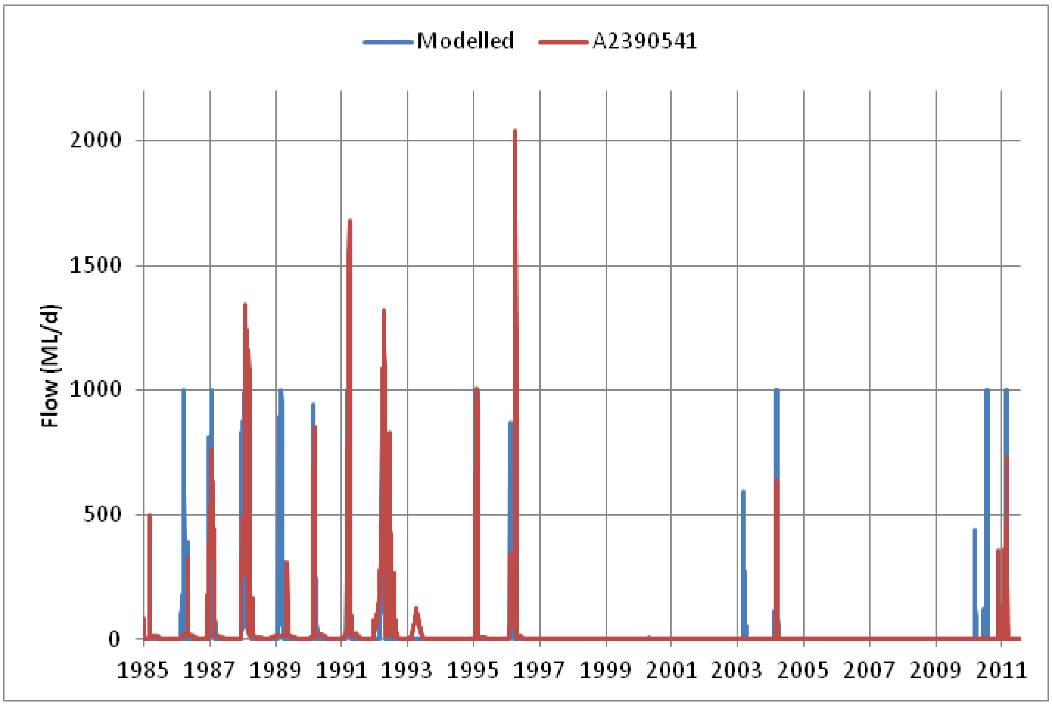


Figure 62 Daily flows downstream of Bool Lagoon

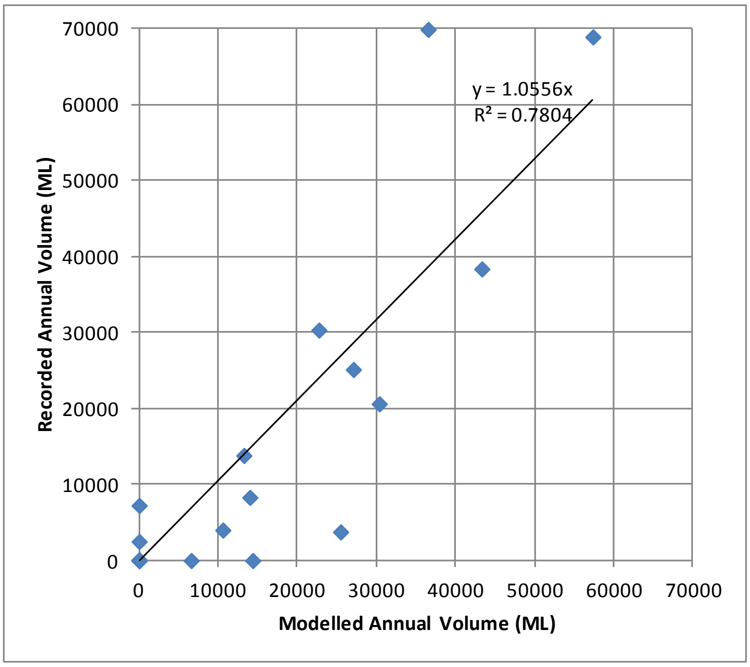


Figure 63 Bool Lagoon recorded and modelled annual release volumes

8.3 Lake George

8.3.1 Dimensions

The Lake George dimensions used were supplied by the South East Water Conservation and Drainage Board, based on surveyed bathymetry undertaken by the Coast Management Branch, and processed by Australian Water Environments (2009).

8.3.2 Climate Data

SILO data at Beachport (26000) has been used for the rainfall and potential evapotranspiration (PET) data for Lake George. The BoM station the SILO data is based on has been open since 1881, and remains open. Class A pan evaporation has been used as the evaporation data, with a pan factor of 0.8 adopted, based on previous studies at Lake George (Australian Water Environments, 2009) and the Coorong (Webster, 2005).

8.3.3 Seepage Rate

Australian Water Environments (2009) adopted a groundwater influx for each month, based on a strip model in an east-west direction through the Upper South East. Australian Water Environments (2009) also suggested the groundwater inflow volumes were thought to be small in comparison to the storage volumes in the lake and hence the model is not sensitive to changes in the groundwater flux. As such, the groundwater influx rates derived by Australian Water Environments (2009) have also been adopted in this work.

8.3.4 Inflow

Inflows to the model include the seepage rate outlined above, the daily flow recorded on Drain M at Woakwine Range (A2390512), and a further 1.1% of this flow to represent Sunderland Drain, which flows into Drain M downstream of the gauge.

There are gates on the outlet of Lake George to the Southern Ocean that can be used to allow flow out of the lake to maintain the target water level, or to allow sea water back into the lake to freshen the salinity during periods of low drain inflow and hypersaline conditions in Lake George. This inflow of sea water has been modelled as the volume required to meet the recorded water level during periods when it was known that the gates were open.

8.3.5 Release Rules

Release rules have been implemented on the Minimum Flow Requirement node downstream of Lake George (Figure 58) to maintain a maximum summer water level threshold (Dec to June) of 0.13 m, and the minimum water level threshold for winter (July to November) of 0.42 m (Department for Environment and Heritage, 2006b). The maximum water level threshold for winter is 0.58 m, however this has not been targeted in the model.

8.3.6 Salinity

A salt balance has also been implemented for Lake George to provide useful information when managing the network, and to also assist in the calibration of the model. The groundwater influx salinity of 467 mg/L used by Australian Water Environments (2009) has also been adopted. A salinity of 35,000 mg/L for seawater inflows has been used. For the salinity of the drain inflow, the line of best fit between daily flow and salt load indicates that the drain flow carries an average salinity of 0.704 t/ML (Figure 64). As such, a salinity of 704 mg/L has been adopted for the drain inflows.

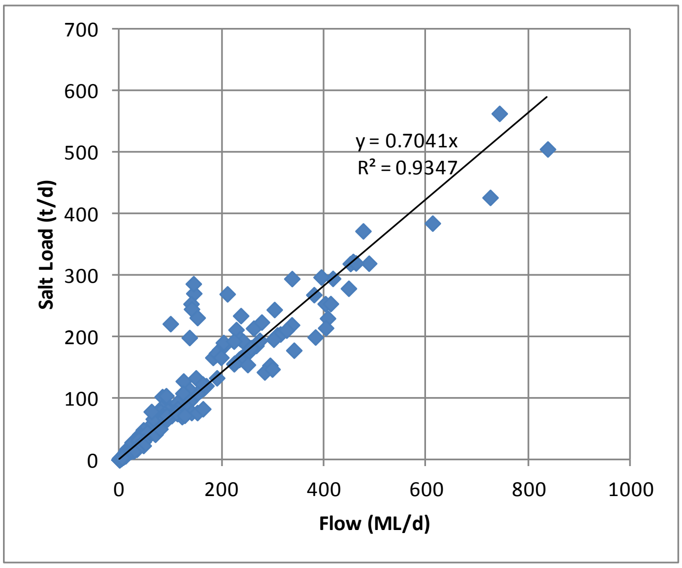


Figure 64 Relationship between salt load and flow at site A2390512

8.3.7 Results

Modelled and observed water levels for Lake George are presented in Figure 65, and salinities in Figure 66. It can be seen from Figure 65 that on a number of occasions the water level in Lake George has exceeded the modelled target level of 0.42 m. However, this is a minimum threshold, as opposed to the maximum threshold of 0.58 m. Generally, the model can be seen to follow the dynamics of the water level fluctuation relatively accurately.

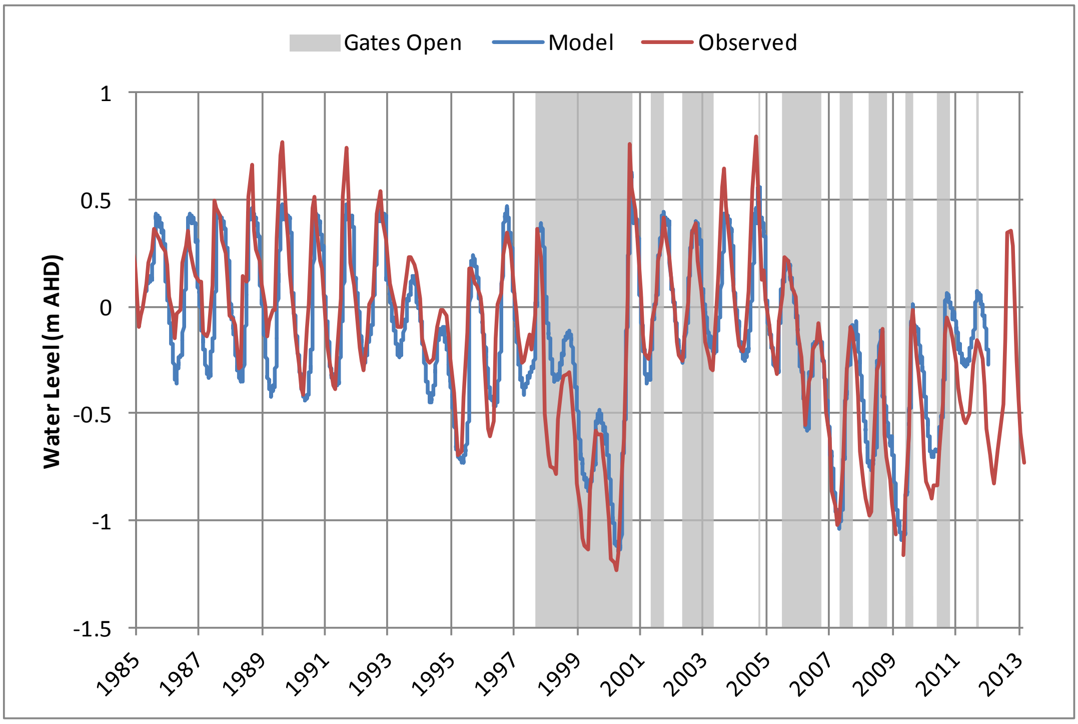


Figure 65 Modelled and observed water level in Lake George

The model does not perform as well for salinity, matching the salinity reasonably well in the early 2000s (regular monthly monitoring of salinity commenced in November 2000), however during the very dry period from 2006 – 2009 the model can be seen to overestimate salinity. This suggests that more sea water flowed into the lake over this period than has been modelled, which would reduce the salinity of the lake to closer to that of sea water (35,000 mg/L). It should also be noted that the salinity sensors tend to under-represent salinity at high salinities, which may also explain some of the difference between the modelled and observed salinity.

Given that a similar pattern of increasing and decreasing salinity was represented in the model it is assumed that the dynamics of the lake are represented reasonably, and with more information on the volume of flow in and out of the connection to the sea, a better representation of the historic salinity data could be obtained. However, this extra input information would not be expected to improve the ability of the model to represent the water and salt balance for Lake George.

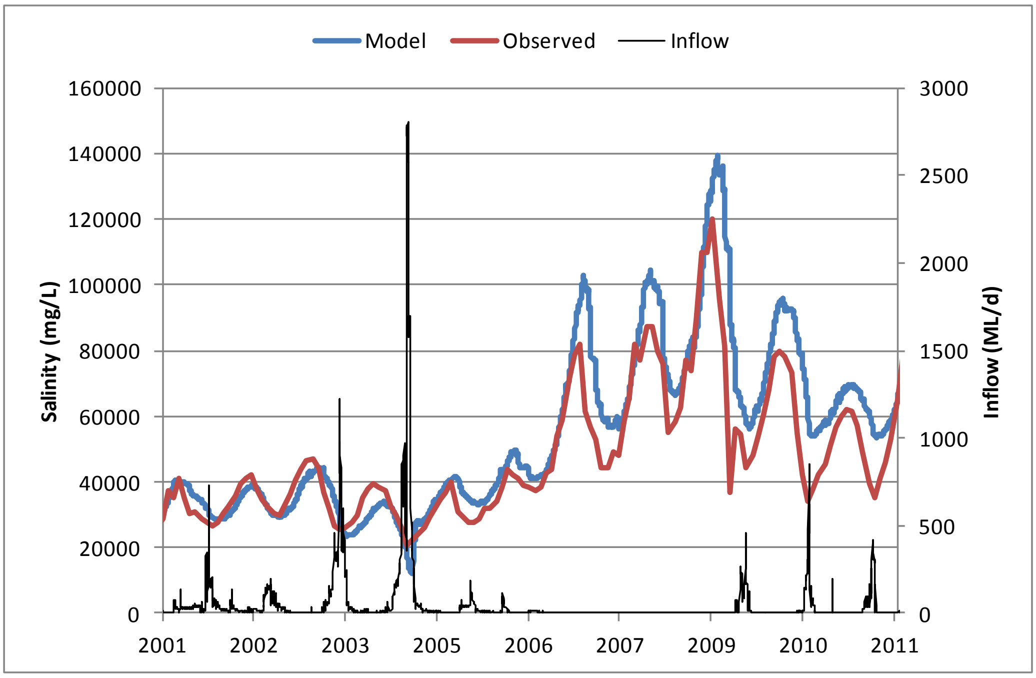


Figure 66 Modelled and observed salinity in Lake George

8.4 Summary

In this section, the development of a Source model to represent the Drain M has been outlined. The model can be seen to provide an accurate representation of the storage and releases Bool and Hacks Lagoon. Considering the Lake George storage is also influenced by inflows from the ocean, the storage in Lake George can also be considered to be represented accurately. The ability to simulate the salinity of Lake George was also included in the model, as this is a key variable in the management of the lake.

The Source model provides the functionality to:

- Take the forecast streamflow volumes in the drainage network from the models developed previously in this report
- Apply different management scenarios for the operation of:
    - Bool Lagoon
    - Callendale regulator and diversions along the REFLOWS drain
    - Lake George
- Assess the:
    - water level in Bool Lagoon
    - water level and salinity in Lake George
    - volume diverted along the REFLOWS drain

This capability is intended to provide greater information to assist the decision making process, extending the probability of a certain volume of water occurring in the network presented in Section 5 to also consider the influence of operational decisions on the key variables (volume diverted, water levels and Lake George salinity) in the system.

9 Conclusions

The objectives of this project were to:

- Assess the potential to include groundwater data as an input to lumped conceptual rainfall-runoff models to improve the ability to simulate changes in the catchments over time (Section 3)
- Develop both lumped conceptual (Section 4) and statistical (Section 5) models to provide predictions of streamflow in Drain M given a seasonal climate forecast
- Assess the accuracy and precision of the different model types, including a combined approach (Section 6)
- Undertake a trial of the streamflow prediction models in the 2014 season (Section 5)
- Develop a water balance model in eWater Source to allow the effect of predicted volumes and different management options on the volumes diverted and water levels in wetlands in the system (Section 8)

The outcomes for each of these objectives are summarised here.

Lumped conceptual Rainfall-Runoff Modelling

Lumped conceptual rainfall-runoff models require some simplification of all the processes that result in streamflow being observed, and as such over time the limitation of the models to represent the streamflow generation processes becomes apparent. Two approaches were adopted to improve the ability of lumped conceptual rainfall runoff models to simulate streamflow.

The first was to vary CRR parameters with groundwater data, to provide another source of information about changes occurring in the catchment for the model. This was found to be beneficial in cases where a change in the rainfall-runoff relationship has been identified, and as such the response at one time was very different to another. Very good agreement between certain model parameter values and the change in catchment averaged groundwater level was found, indicating that there is the potential to inform time varying conceptual rainfall runoff model parameters using groundwater data. However, a threshold behaviour in the model parameter values that co-varied with the groundwater trend information was observed, where one value was most appropriate under one set of conditions, and a second value after a change in groundwater level of a certain magnitude occurred. As such, for seasonal forecasting, calibrating the model parameters to simulate the most recent data, representing the current catchment conditions, may be a more suitable approach to determine the model parameters, without the unnecessary complication of introducing further input data requirements and increasing model complexity. For longer time simulations, such as for water allocation planning, accounting for trends in the catchment, through changes in groundwater level or otherwise, is likely to be a beneficial approach.

The second approach involved assessing approaches to estimate the total predictive uncertainty when applying CRR for monthly streamflow predictions. The predictive uncertainty was estimated by quantifying the parameter uncertainty, and estimation of the residual error through a post processing approach. A number of methods were tested to provide the most reliable and precise estimate of the predictive uncertainty as possible. It was found that a log transformation of the residual errors allowed the heteroscedasticity in model errors to be accounted for in the most precise way, and assimilation of the observed flow available prior to making a forecast allowed for the most reliable correction to any biases in the model outputs. Two likelihood functions selected to quantify how representative a parameter set was of the catchment given the data available were also tested, with a standard sum of squared errors based function providing the most accurate representation of high flow events, with the highest precision in the predictions as well. Based on these investigations, a number of improvements have been made to the ability to simulate streamflow, and estimate the uncertainty involved, in the South East of South Australia with lumped conceptual rainfall-runoff models.

Statistical Models

In order to provide forecasts of future streamflows, statistical models require predictors that describe the initial catchment condition and the effect of climate during the forecast period. Predictions of the future climate were provided by the Bureau of Meteorology's POAMA forecast system. As well as traditional predictors such as previous streamflows, rainfall and evapotranspiration, a number of datasets were also considered to describe the initial catchment conditions:

- A remotely sensed soil moisture data set, produced by the European Space Agency combining both passive and actively sensed soil moisture measurements
- A modelled soil moisture from the coupled ocean/atmosphere POAMA model used to provide the climate forecasts
- The soil storage level modelled by the CRRs

Each of the soil moisture datasets considered provide different information about soil water storage and have their own advantages and limitations. Overall, it was found that the inclusion of all three datasets was required to give the best improvement in forecast accuracy and reliability over the model which only included more traditional inputs for representing antecedent catchment conditions and did not include soil moisture data directly. It was also found that each individual soil moisture dataset was able to provide some improvement over not including any soil moisture data, particularly in terms of mean forecast accuracy.

Based on these results, when developing statistical streamflow forecasting models, it is considered worthwhile to at least include soil moisture data obtained from large, readily available global datasets (such as POAMA SM and ECV SM) as potential model inputs. It is not recommended, however, that soil moisture data be included at the expense of more traditional types of inputs used for characterising initial catchment conditions, since the results of the input variable selection performed found that such inputs, including previous streamflows and rainfall, are also necessary for making accurate streamflow forecasts. This result highlights the importance of adopting a rigorous input variable selection method to select those inputs that are most relevant to the problem at hand.

Model Comparison

A number of different models were developed to predict monthly streamflow in Drain M. The performance of the different models was compared, both in terms of the accuracy of the mean model outputs, as well as the ability to capture the uncertainty range in flow. The comparison was made between: 1) the lumped conceptual models, 2) the statistical models, and 3) a combined model, where the outputs from the conceptual models was available as an input to the statistical models.

It was found that that the combined model produced robust results across the three catchments, with all observations within the modelled forecast limits, and low relative confidence intervals indicating that the estimated forecast uncertainty was not unnecessarily large. This can be contrasted to the conceptual model alone, which provided comparable predictions for one catchment, but much wider confidence bounds for the another catchment, and less satisfactory results on the third catchment considered.

The advantages of the statistical/combined models was 1) their ability to generalise the relationship of interest, and as such provide the good performance consistently across the three catchments considered and 2) their ability to make use of a range of different data sets available that were found to be related to streamflow. The main disadvantage of these models is related to this second benefit, where increased ability to make use of different data sources can increase the number and range of inputs that must be collated to calculate the model output of interest (i.e. streamflow). For one catchment this increase in complexity did not substantially improve the predictions from the model. In contrast the advantages of the lumped conceptual models are their simplicity, where for one catchment comparable results could be obtained from only readily available climate information. However, the disadvantage of these models is they must be prescribed *a priori*, and the simplified form of the model may not be appropriate for the catchment of interest. As such, much more care is required in selecting a suitable lumped conceptual model on a case by case basis.

2014 Forecast Trial

The lumped conceptual models were used to forecast the flow in Drain M one month in advance during the 2014 season. The lumped conceptual models were used as the remotely sensed soil moisture data selected as an input for the combined model is not currently available in real time. However, this data is likely to become more readily available in the future, with missions such as the NASA Soil Moisture Active Passive (SMAP) satellite recently launched.

The value of information on antecedent conditions is clear from this trial, where the assimilation of previously observed data and warm up of the lumped conceptual model using recent climate data improve the reliability of the forecasts compared to the rainfall forecast, particularly for the extremely dry months experienced.

The models estimated well the observed flow events that occurred based on observed rainfall, indicating that the models developed are likely to have some use in informing the management of the drainage network. However, the flow at the start of the season was overestimated, and a different model structure could be considered to have a better representation of the commence-to-flow for the season, and low flows more generally.

Over the long term, the downscaled POAMA rainfall forecasts were found to be reliable. However, the monthly rainfall was substantially overestimated in some months during the 2014 trial. While incorporating the influence of antecedent conditions in the catchment has been demonstrated to improve the results, the accuracy of streamflow forecasts are directly linked to the accuracy of rainfall forecasts.

Water Balance model

An operational model has also been developed to allow for different diversion scenarios to be assessed. The model was found to provide an accurate representation of the storage and releases from Bool and Hacks Lagoon as well as the storage and salinity in Lake George.

The Source model provides the functionality to:

- Take the forecast streamflow volumes in the drainage network from the models developed
- Apply different management scenarios for the operation of:
    - Bool Lagoon
    - Callendale regulator and diversions along the REFLOWS drain
    - Lake George
- Assess the:
    - water level in Bool Lagoon
    - water level and salinity in Lake George
    - volume diverted along the REFLOWS drain

This capability is intended to provide greater information to assist the decision making process, extending the probability of a certain volume of water occurring in the network from the forecast models, to also consider the influence of operational decisions on the key variables (volume diverted, water levels and Lake George salinity) in the system.

10 Appendix A

Candidate inputs considered for each catchment for the statistical models with and without the conceptual model (GR4J) soil moisture input.

Table A.1 Candidate inputs for the additional model developed to forecast flow at gauge A2390512. See Table 7 for the candidate inputs for the models developed in Section 5 to forecast flow at A2390512.

base+ECV SM+POAMA SM: – A total of 11 candidate inputs: • 7 ***base*** model inputs; • 2 additional ***base + ECV SM*** model inputs (see Table 9); and • 2 additional ***base + POAMA SM*** model inputs (see Table 9)

Table A.2 Candidate inputs for the models developed to forecast flow at gauge A2390514

base model***:*** a total of 39 candidate inputs:

- Flow_A2390514$_{t-i}$, Flow_A2390541$_{t-i}$, Flow_A2390519$_{t-i}$,
- Rain_A2390514$_{t-i}$, Rain_A2390519$_{t-i}$,
- Evap_A2390514$_{t-i}$, Evap_A2390519$_{t-i}$,
- API_A2390514$_{t-i}$, API_A2390519$_{t-i}$,
- GW_CMM079$_{t-i}$,
- POAMA_Rain_26003_min$_{t+1}$, POAMA_Rain_26003_mean$_{t+1}$, POAMA_Rain_26003_max$_{t+1}$;
- POAMA_Rain_26075_min$_{t+1}$, POAMA_Rain_26075_mean$_{t+1}$, POAMA_Rain_26075_max$_{t+1}$;
- POAMA_Rain_26082_min$_{t+1}$, POAMA_Rain_26082_mean$_{t+1}$, POAMA_Rain_26082_max$_{t+1}$;

where i = 0, 1, 2

base + ECV SM+POAMA SM: a total of 68 candidate inputs:

- 8 ***base*** model inputs (see Table 14),
- ECV_SM_A2390514$_{t-i}$,
- ECV_SM_A2390515$_{t-i}$,
- ECV_SM_A2391001$_{t-i}$,
- ECV_SM_DrainC$_{t-i}$,
- ECV_SM_A2390516$_{t-i}$,
- ECV_SM_A2390537$_{t-i}$,
- ECV_SM_A2390536$_{t-i}$,
- ECV_SM_A2390541$_{t-i}$,
- ECV_SM_A2390519$_{t-i}$,
- ECV_SM_A2391076$_{t-i}$;
- POAMA_SM_A2390514_min$_{t+1}$, POAMA_SM_A2390514_mean$_{t+1}$, POAMA_SM_A2390514_max$_{t+1}$,
- POAMA_SM_A2390515_min$_{t+1}$, POAMA_SM_A2390515_mean$_{t+1}$, POAMA_SM_A2390515_max$_{t+1}$,
- POAMA_SM_A2391001_min$_{t+1}$, POAMA_SM_A2391001_mean$_{t+1}$, POAMA_SM_A2391001_max$_{t+1}$,
- POAMA_SM_DrainC_min$_{t+1}$, POAMA_SM_DrainC_mean$_{t+1}$, POAMA_SM_DrainC_max$_{t+1}$,
- POAMA_SM_A2390516_min$_{t+1}$, POAMA_SM_A2390516_mean$_{t+1}$, POAMA_SM_A2390516_max$_{t+1}$,
- POAMA_SM_A2390537_min$_{t+1}$, POAMA_SM_A2390537_mean$_{t+1}$, POAMA_SM_A2390537_max$_{t+1}$,
- POAMA_SM_A2390536_min$_{t+1}$, POAMA_SM_A2390536_mean$_{t+1}$, POAMA_SM_A2390536_max$_{t+1}$,
- POAMA_SM_A2390541_min$_{t+1}$, POAMA_SM_A2390541_mean$_{t+1}$, POAMA_SM_A2390541_max$_{t+1}$,
- POAMA_SM_A2390519_min$_{t+1}$, POAMA_SM_A2390519_mean$_{t+1}$, POAMA_SM_A2390519_max$_{t+1}$,
- POAMA_SM_A2391076_min$_{t+1}$, POAMA_SM_A2391076_mean$_{t+1}$, POAMA_SM_A2391076_max$_{t+1}$

where i = 0, 1, 2

base + GR4J SM: a total of 14 candidate inputs:

- 8 ***base*** model inputs,
- GR4J_SM_A2390514$_{t-i}$;
- GR4J_SM_A2390519$_{t-i}$;

where i = 0, 1, 2

base + all SM: – A total of 13 candidate inputs:

- 8 ***base*** model inputs;
- 1 additional ***base + ECV SM*** model input (see Table 14);
- 2 additional ***base + POAMA SM*** model inputs (see Table 14); and
- 2 additional ***base + GR4J SM*** model inputs (see Table 14).

Table A.3 Candidate inputs for the models developed to forecast flow at gauge A2390519

base model***:*** a total of 18 candidate inputs:

- Flow_A2390519$_{t-i}$,
- Rain_A2390519$_{t-i,}$
- Evap_A2390519$_{t-i}$,
- API_A2390519$_{t-i}$,
- POAMA_Rain_26075_min$_{t+1}$, POAMA_Rain_26075_mean$_{t+1}$, POAMA_Rain_26075_max$_{t+1}$;
- POAMA_Rain_26082_min$_{t+1}$, POAMA_Rain_26082_mean$_{t+1}$, POAMA_Rain_26082_max$_{t+1}$;

where i = 0, 1, 2

base + ECV SM+POAMA SM: a total of 17 candidate inputs:

- 5 ***base*** model inputs (see Table 15),
- ECV_SM_A2390519$_{t-i}$,
- ECV_SM_A2391076$_{t-i}$;
- POAMA_SM_A2390519_min$_{t+1}$, POAMA_SM_A2390519_mean$_{t+1}$, POAMA_SM_A2390519_max$_{t+1}$,
- POAMA_SM_A2391076_min$_{t+1}$, POAMA_SM_A2391076_mean$_{t+1}$, POAMA_SM_A2391076_max$_{t+1}$

where i = 0, 1, 2

base + GR4J SM: a total of 8 candidate inputs:

- 5 ***base*** model inputs,
- GR4J_SM_A2390519$_{t-i}$;

where i = 0, 1, 2

base + all SM: – A total of 10 candidate inputs:

- 5 ***base*** model inputs;
- 4 additional ***base + ECV SM*** model inputs (see Table 15);
- 1 additional ***base + POAMA SM*** model input (see Table 15); and
- 1 additional ***base + GR4J SM*** model input (see Table 15).

11 References

Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998. Crop evapotranspiration-Guidelines for computing crop water requirements. FAO Irrigation and drainage paper 56, FAO, Rome.

Anctil, F., Lauzon, N., Filiona, M., 2008. Added gains of soil moisture content observations for streamflow predictions using neural networks. Journal of Hydrology, Volume 359(3–4): 225–234.

Anctil, F., Michel, C., Perrin, C., Andr\'eassian, V., 2004. A soil moisture index as an auxiliary ANN input for stream flow forecasting. J. Hydrol., 286(1–4): 155–167.

Andrews, F.T., Croke, B.F.W., Jakeman, A.J., 2011. An open software environment for hydrological model assessment and development. Environmental Modelling & Software, 26(10): 1171-1185.

Australian Water Environments, 2009. Lake George Salt and Water Balance.

Bennett, J.C., Wang, Q.J., Pokhrel, P., Robertson, D.E., 2014. The challenge of forecasting high streamflows 1–3 months in advance with lagged climate indices in southeast Australia. Nat. Hazards Earth Syst. Sci., 14(2): 219-233.

Berg, A.A., Mulroy, K.A., 2006. Streamflow predictability in the {S}askatchewan{/N}elson River basin given macroscale estimates of the initial soil moisture status. Hydrological Sciences Journal, 51(4): 642-654.

Berthet, L., Andreassian, V., Perrin, C., Javelle, P., 2009. How crucial is it to account for the antecedent moisture conditions in flood forecasting? Comparison of event-based and continuous approaches on 178 catchments. Hydrology and Earth System Sciences, 13(6): 819-831.

Beven, K.J., Smith, P.J., Freer, J.E., 2008. So just why would a modeller choose to be incoherent? J. Hydrol., 354(1–4): 15-32.

Bindlish, R., Crow, W.T., Jackson, T.J., 2009. Role of passive microwave remote sensing in improving flood forecasts. Geoscience and Remote Sensing Letters, IEEE, 6(1): 112-116.

Blum, A.L., Langley, P., 1997. Selection of relevant features and examples in machine learning. Artificial Intelligence, 97(1–2): 245-271.

Bowden, G.J., Maier, H.R., Dandy, G.C., 2012. Real-time deployment of artificial neural network forecasting models: Understanding the range of applicability. Water Resour. Res., 48(10): n/a-n/a.

Brigode, P., Oudin, L., Perrin, C., 2013. Hydrological model parameter instability: A source of additional uncertainty in estimating the hydrological impacts of climate change? J. Hydrol., 476: 410-425.

Brocca, L., Melone, F., Moramarco, T., 2008. On the estimation of antecedent wetness conditions in rainfall--runoff modelling. Hydrological Processes, 22: 629–642.

Brocca, L. et al., 2012. Assimilation of Surface- and Root-Zone ASCAT Soil Moisture Products Into Rainfall–Runoff Modeling. Geoscience and Remote Sensing, IEEE Transactions on, 50(7): 2542-2555.

Casper, M., Gemmar, P., Gronz, O., Johst, M., Stüber, M., 2007. Fuzzy logic-based rainfall—runoff modelling using soil moisture measurements to represent system state. Hydrological Sciences Journal, 52(3): 478-490.

Cheng, Q.-B., Chen, X., Xu, C.-Y., Reinhardt-Imjela, C., Schulte, A., 2014. Improvement and comparison of likelihood functions for model calibration and parameter uncertainty analysis within a Markov chain Monte Carlo scheme. J. Hydrol., 519, Part B(0): 2202-2214.

Coron, L. et al., 2012. Crash testing hydrological models in contrasted climate conditions: An experiment on 216 Australian catchments. Water Resour. Res., 48: W05552.

Dash, M., Liu, H., 1997. Feature selection for classification. Intelligent Data Analysis, 1(3): 131-156.

Dawson, C.W., Abrahart, R.J., See, L.M., 2007. HydroTest: A web-based toolbox of evaluation metrics for the standardised assessment of hydrological forecasts. Environmental Modelling \& Software, 22(7): 1034-1052.

De Jeu, R.A.M. et al., 2012. State of the climate in 2011: Sidebar 2.2: Building a climate record of soil moisture from historical satellite observations. Bulletin of the American Meteorological Society, 93(7): S32-S33.

Demargne, J. et al., 2013. The Science of NOAA's Operational Hydrologic Ensemble Forecast Service. Bulletin of the American Meteorological Society, 95(1): 79-98.

Demirel, M.C., Booij, M.J., Hoekstra, A.Y., 2013. Effect of different uncertainty sources on the skill of 10 day ensemble low flow forecasts for two hydrological models. Water Resour. Res., 49(7): 4035-4053.

Department for Environment and Heritage, 2006a. Bool Lagoon Game Reserve and Hacks Lagoon Conservation Park Management and Ramsar Plan, Adelaide, South Australia.

Department for Environment and Heritage, 2006b. Lake George Management Plan, Department for Environment and Heritage and the Lake George Management Committee.

Dorigo, W. et al., 2012. Evaluating global trends (1988–2010) in harmonized multi-satellite surface soil moisture. Geophysical Research Letters, 39(18): n/a-n/a.

Draper, C.S., Walker, J.P., Steinle, P.J., de Jeu, R.A.M., Holmes, T.R.H., 2009. An evaluation of AMSR–E derived soil moisture over Australia. Remote Sensing of Environment, 113(4): 703-710.

Duan, Q., Sorooshian, S., Gupta, V.K., 1994a. Optimal use of the SCE-UA global optimization method for calibrating watershed models. J. Hydrol., 158(3–4): 265-284.

Duan, Q.Y., Sorooshian, S., Gupta, V.K., 1994b. optimal use of the SCE-UA global optimization method for calibrating watershed models. J. Hydrol., 158(3-4): 265-284.

Evin, G., Thyer, M., Kavetski, D., McInerney, D., Kuczera, G., 2014. Comparison of joint versus postprocessor approaches for hydrological uncertainty estimation accounting for error autocorrelation and heteroscedasticity. Water Resour. Res., 50(3): 2350-2375.

Ferdowsian, R., Pannell, D.J., McCarron, C., Ryder, A., Crossing, L., 2001. Explaining groundwater hydrographs: separating atypical rainfall events from time trends. Soil Research, 39(4): 861-876.

Galelli, S. et al., 2014. An evaluation framework for input variable selection algorithms for environmental data-driven models. Environmental Modelling & Software, 62(0): 33-51.

Gautam, M.R., Watanabe, K., Saegusa, H., 2000. Runoff analysis in humid forest catchment with artificial neural network. J. Hydrol., 235(1–2): 117-136.

Goldberg, D.E., 1989. Genetic Algorithms in Search, Optimization and Machine Learning. Addison-Wesley Pub. Co., Reading, Mass.

Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. J. Hydrol., 377(1–2): 80-91.

Guyon, I., Elisseeff, A.e., 2003. An introduction to variable and feature selection. The Journal of Machine Learning Research, 3: 1157–1182.

Haario, H., Saksman, E., Tamminen, J., 2001. An adaptive Metropolis algorithm. Bernoulli, 7(2): 223-242.

Heggen, R., 2001. Normalized Antecedent Precipitation Index. Journal of Hydrologic Engineering, 6(5): 377-381.

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

Jeffrey, S.J., Carter, J.O., Moodie, K.B., Beswick, A.R., 2001. Using spatial interpolation to construct a comprehensive archive of Australian climate data. Environmental Modelling & Software, 16(4): 309-330.

Kavetski, D., Franks, S.W., Kuczera, G., 2003. Confronting Input Uncertainty in Environmental Modelling, Calibration of Watershed Models. American Geophysical Union, pp. 49-68.

Kingston, G.B., Lambert, M.F., Maier, H.R., 2005. Bayesian training of artificial neural networks used for water resources modeling. Water Resources Research, 41(12): W12409.

Kokkonen, T.S., Jakeman, A.J., 2001. A comparison of metric and conceptual approaches in rainfall-runoff modeling and its implications. Water Resour. Res., 37(9): 2345-2352.

Koren, V., Moreda, F., Smith, M., 2008. Use of soil moisture observations to improve parameter consistency in watershed calibration. Physics and Chemistry of the Earth, Parts A/B/C, 33(17–18): 1068-1080.

Kuczera, G., 1983. Improved parameter inference in catchment models: 1. Evaluating parameter uncertainty. Water Resources Research, 19(5): 1151-1162.

Lavery, B., Kariko, A., Nicholls, N., 1992. A historical rainfall data set for Australia. Australian Meteorological Magazine, 40: 33-39.

Le Moine, N., Andreassian, V., Perrin, C., Michel, C., 2007. How can rainfall-runoff models handle intercatchment groundwater flows? Theoretical study based on 1040 French catchments. Water Resour. Res., 43(6): W06428.

Linares-Rodriguez, A., Lara-Fanego, V., Pozo-Vazquez, D., Tovar-Pescador, J., 2015. One-day-ahead streamflow forecasting using artificial neural networks and a meteorological mesoscale model. Journal of Hydrologic Engineering, 0(0): 05015001.

Liu, H., Motoda, H., Setiono, R., Zhao, Z., 2010. Feature selection: an ever evolving frontier in data mining the 4th International Workshop on Feature Selection in Data Mining, Hyderabad, India, pp. 4-13.

Luo, J., Wang, E., Shen, S., Zheng, H., Zhang, Y., 2012. Effects of conditional parameterization on performance of rainfall-runoff model regarding hydrologic non-stationarity. Hydrological Processes, 26(26): 3953-3961.

Maier, H.R., Dandy, G.C., 2000. Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. Environmental Modelling & Software, 15(1): 101-124.

Maier, H.R., Jain, A., Dandy, G.C., Sudheer, K.P., 2010. Methods used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future directions. Environmental Modelling & Software, 25(8): 891-909.

Manabe, S., Holloway, J.L., 1975. The seasonal variation of the hydrologic cycle as simulated by a global model of the atmosphere. Journal of Geophysical Research, 80(12): 1617-1649.

May, R.J., Dandy, G.C., Maier, H.R., 2011. Review of input variable selection methods for artificial neural networks. In: Suzuki, K. (Ed.), Artificial Neural Networks - Methodological Advances and Biomedical Applications. InTech, Rijeka, Croatia.

Morawietz, M., Xu, C.-Y., Gottschalk, L., Tallaksen, L.M., 2011. Systematic evaluation of autoregressive error models as post-processors for a probabilistic streamflow forecast system. Journal of Hydrology, 407(1–4): 58-72.

Moriasi, D.N. et al., 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Transactions of the ASABE, 50(3): 885−900.

Mullen, K.M., Ardia, D., Gil, D.L., Windover, D., Cline, J., 2010. DEoptim: An R Package for Global Optimization by Differential Evolution, University Library of Munich, Germany.

Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I - a discussion of principles. J. Hydrol., 10(3): 282-290.

Perrin, C., Michel, C., Andreassian, V., 2003. Improvement of a parsimonious model for streamflow simulation. J. Hydrol., 279(1-4): 275-289.

Petheram, C., Potter, N., Vaze, J., Chiew, F., Zhang, L., 2011. Towards better understanding of changes in rainfall-runoff relationships during the recent drought in south-eastern Australia, MODSIM 2011, 19th International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand, Perth, Australia.

Pokhrel, P., Wang, Q.J., Robertson, D.E., 2012. Merging statistical and dynamical forecasting models to improve seasonal streamflow forecasts in eastern Australian catchments, 34th Hydrology and Water Resources Symposium Engineers Australia, Sydney, Australia, pp. 1173-1180.

Renard, B., Kavetski, D., Kuczera, G., Thyer, M., Franks, S.W., 2010. Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors. Water Resources Research, 46.

Robertson, D.E., Pokhrel, P., Wang, Q.J., 2013. Improving statistical forecasts of seasonal streamflows using hydrological model output. Hydrology and Earth System Sciences, 17(2): 579–593.

Robertson, D.E., Wang, Q.J., 2012. A Bayesian approach to predictor selection for seasonal streamflow forecasting. Journal of Hydrometeorology, 13(1): 155-171.

Robertson, D.E., Wang, Q.J., 2013. Seasonal Forecasts of Unregulated Inflows into the Murray River, Australia. Water Resour Manage, 27(8): 2747-2769.

Schoups, G., Vrugt, J.A., 2010. A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors. Water Resources Research, 46.

Scipal, K., Scheffler, C., Wagner, W., 2005. Soil moisture-runoff relation at the catchment scale as observed with coarse resolution microwave remote sensing. Hydrology and Earth System Sciences, 9(3): 173-183.

Shamseldin, A.Y., 1997. Application of a neural network technique to rainfall-runoff modelling. J. Hydrol., 199(3–4): 272-294.

Shao, Q., Li, M., 2013. An improved statistical analogue downscaling procedure for seasonal precipitation forecast. Stoch Environ Res Risk Assess, 27(4): 819-830.

Sorooshian, S., Dracup, J.A., 1980. Stochastic parameter estimation procedures for hydrologie rainfall-runoff models: Correlated and heteroscedastic error cases. Water Resources Research, 16(2): 430-442.

Tayfur, G., Zucco, G., Brocca, L., Moramarco, T., 2014. Coupling soil moisture and precipitation observations for predicting hourly runoff at small catchment scale. J. Hydrol., 510: 363-371.

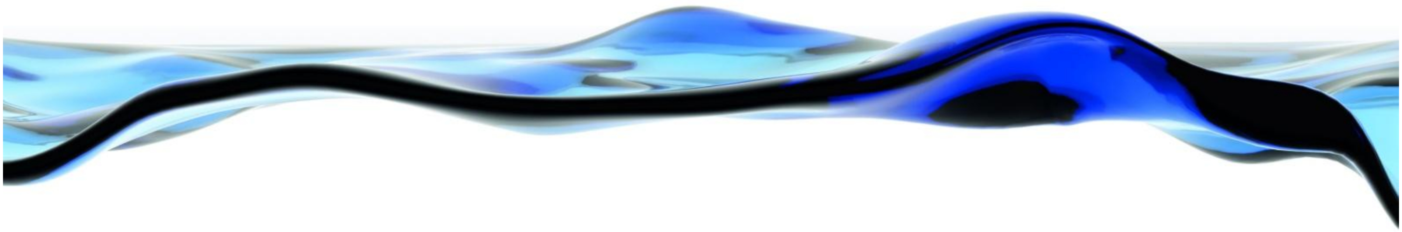
Thyer, M. et al., 2009. Critical evaluation of parameter consistency and predictive uncertainty in hydrological modeling: A case study using Bayesian total error analysis. Water Resources Research, 45.

Tolson, B.A., Shoemaker, C.A., 2007. Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. Water Resour. Res., 43(1): W01413.

Tomkins, K.M., 2014. Uncertainty in streamflow rating curves: methods, controls and consequences. Hydrological Processes, 28(3): 464-481.

Toth, E., Brath, A., 2002. Flood forecasting using artificial neural networks in black-box and conceptual rainfall-runoff modelling. In: Rizzoli, A.E., Jakeman, A.J. (Eds.), Integrated Assessment and Decision Support, Proceedings of the 1st Biennial Meeting of the International Environmental Modelling and Software Society. iEMSs, Lugano, Switzerland, pp. 166-171.

Trewin, B., 2013. A daily homogenized temperature data set for Australia. International Journal of Climatology, 33(6): 1510-1529.
Vaze, J. et al., 2010. Climate non-stationarity – Validity of calibrated rainfall–runoff models for use in climate change studies. J. Hydrol., 394(3–4): 447-457.
Viney, N.R. et al., 2009. The usefulness of bias constraints in model calibration for regionalisation to ungauged catchments, 18th World IMACS / MODSIM Congress, Cairns, Australia, pp. 1-7.
Vrugt, J.A., Diks, C.G.H., Gupta, H.V., Bouten, W., Verstraten, J.M., 2005. Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation. Water Resources Research, 41(1).
Vrugt, J.A. et al., 2009. Accelerating Markov Chain Monte Carlo Simulation by Differential Evolution with Self-Adaptive Randomized Subspace Sampling. International Journal of Nonlinear Sciences and Numerical Simulation, 10(3): 273-290.
Wang, E., Zhang, Y., Luo, J., Chiew, F.H.S., Wang, Q.J., 2011. Monthly and seasonal streamflow forecasts using rainfall-runoff modeling and historical weather data. Water Resour. Res., 47(5): W05516.
Webster, I.T., 2005. An Overview of the Hydrodynamics of the Coorong and Murray Mouth. Technical Report No. #/2005, CSIRO: Water for a Healthy Country National Research Flagship.
Węglarczyk, S., 1998. The interdependence and applicability of some statistical quality measures for hydrological models. J. Hydrol., 206(1–2): 98-103.
Westenbroek, S.M. et al., 2012. Approaches in Highly Parameterized Inversion: TSPROC, a general time-series processor to assist in model calibration and result summarization, USGS.
White, J.T., Doherty, J.E., Hughes, J.D., 2014. Quantifying the predictive consequences of model error with linear subspace analysis. Water Resour. Res., 50(2): 1152-1173.
Wijngaard, J.B., Klein Tank, A.M.G., Können, G.P., 2003. Homogeneity of 20th century European daily temperature and precipitation series. International Journal of Climatology, 23(6): 679-692.
Wood, A.W., Schaake, J.C., 2008. Correcting Errors in Streamflow Forecast Ensemble Mean and Spread. Journal of Hydrometeorology, 9(1): 132-148.
Wood, G., Way, D., 2011. Development of the Technical Basis for a Regional Flow Management Strategy for the South East of South Australia, Adelaide.
Wooldridge, S.A., Kalma, J.D., Walker, J.P., 2003. Importance of soil moisture measurements for inferring parameters in hydrologic models of low-yielding ephemeral catchments. Environmental Modelling & Software, 18(1): 35–48.
Xu, X., Li, J., Tolson, B.A., 2014. Progress in integrating remote sensing data and hydrologic modeling. Progress in Physical Geography, 38(4): 464-498.



The Goyder Institute for Water Research is a partnership between the South Australian Government through the Department of Environment, Water and Natural Resources, CSIRO, Flinders University, the University of Adelaide and the University of South Australia.