1 **From gridded runoff to streamflow: application of statistical post-processing to generate seasonal streamflow**

2 **forecasts.**

3 Christopher A. Pickett-Heaps[1], Patrick Sunter [2], Wendy Sharples [2], Michael Pegios [2], Catherine

4 Wilson [2], Alex Cornish [3], Richard Laugesen [1] & Elisabetta Carrara [2]

5 [1] Bureau of Meteorology, Canberra, Australia

6 [2] Bureau of Meteorology, Melbourne, Australia

7 [3] Bureau of Meteorology, Adelaide, Australia

8 Correspondence to: Christopher A. Pickett-Heaps (christopher.pickett-heaps@bom.gov.au)

9 **Abstract**

10 Hydrological models can be categorized as either fully distributed grid-based models or catchment-based models

11 coincident with river gauging stations. Grid-based models provide national coverage at a relatively high spatial

12 resolution. Catchment-based models target specific catchments upstream of a river-gauging station. While not

13 providing seamless national coverage, catchment-specific model calibration allows catchment-based models to

14 achieve improved performance relative to a distributed model using a single optimized parameter set. Catchment-

15 based models often rely on the supply of real-time hydrological observations, the latency of which hinders the

16 efficiency and timelessness of forecast generation and publication. This study evaluates whether a nation-wide grid-

17 based hydrological model, coupled with statistical post-processing, generates comparable forecast skill to that of a

18 catchment-based statistical hydrological model.

19

20 Two hydrological models are evaluated at 449 gauging stations across Australia. The Australian Water Resource and

21 Assessment model (AWRA-L) is a 5km2 distributed landscape model that provides seasonal forecasts of key

22 hydrological variables (soil-moisture and runoff) at the monthly temporal resolution. The necessary climate forcing is

23 provided by the ACCESS-S seasonal climate forecasting model. Coupled with forecast post-processing, seasonal

24 forecasts of river discharge (streamflow) can be generated from AWRA-L seasonal forecasts at gauged catchments

25 across Australia. The second model evaluated in this study is a statistical hydrological model that is calibrated locally

26 at the same individual catchments, where observed catchment conditions (antecedent observed streamflow) is the

27 primary input variable with which to forecast streamflow.

28

29 Results from this study indicate that seasonal forecasts from a distributed hydrological model coupled with statistical

30 post-processing achieve similar forecast skill to a locally calibrated statistical hydrological model where the observed

31 catchment conditions represented by observed antecedent streamflow is the only input variable. The statistical post-

32 processor is calibrated against historical streamflow observations but excludes antecedent observed streamflow as an

33 input variable. For many catchments, AWRA-L can represent initial catchment conditions as well as the hydrological

34 response arising from climate forcing from a seasonal climate model. The post-processing model provides a bias-

35 correction that is a function of the magnitude of a hydrological model response only (e.g. runoff).

Forecast performance of the AWRA-L model is highest in the high-flow season across Australia but is limited in the low-flow season in certain regions (e.g. the tropical dry season in northern Australia). Inclusion of antecedent streamflow observations as an input variable results in excellent performance due to strong hydrological persistence in low flows. Depending on the location and time of year, either root-zone soil-moisture or runoff are found to maximise forecast skill. Root-zone soil-moisture maximises forecast skill for a higher proportion of sites than runoff overall, particularly in the low-flow season. As forecast skill increases, the proportion of sites for which runoff maximises forecast skill increases, particularly in the high-flow season. Results also indicate the degree to which climate forcing from a seasonal climate model contributes to forecast skill, as distinct to initial catchment conditions and hydrological persistence.

## 1. Introduction

Australia is impacted by a highly variable climate. Multiple global climate drivers (Risbey et al., 2009) influence this variability over different timescales – The El-Nino Southern Oscillation (ENSO, Chiew et al., 1998), the Indian Ocean Dipole (IOD, Ashok et al., 2003), the Southern Annular Mode (SAM, Hendon et al., 2007) and the Madden-Julian Oscillation (MJO, Wheeler et al., 2009). On occasion they act in concert (Ummenhofer et al., 2011), further enhancing climate extremes. A recent example was the severe drought of 2017-2020 (Devanand et al., 2024; Holgate et al., 2020) culminating in the 'Black-Summer Bushfires (Boer et al., 2020), followed by a rare triple La Nina event of 2020-2023 (Huang et al., 2024). In the space of seven years, Australia experienced one of the most intense drought conditions and bushfire season in the modern era, followed by significant wide-spread flood events. Twenty years earlier – a prolonged 10-year drought of the 2000s (Van Dijk, Beck, et al., 2013) preceded an intense double La Nina event (Beard et al. 2011). The La Nina event resulted in a 5mm fall in global mean sea-level, a proportion of which was directly linked to widespread flooding across Australia (Boening et al. 2011).

Hydrological impacts arising from this climate variability negatively impact our urban and natural environments, as well as the socio-economic activities on which our communities depend. Effective and proactive water resource management is needed to mitigate these impacts. Vulnerability to hydrological extremes and limited water resources has led to an investment in seasonal climate and hydrological forecasting services to support water management in Australia. The Australian Bureau of Meteorology[1] (herein referred to as the Bureau) provides two such services to the Australian community: The Australian Water Outlook [2](AWO) and the Seasonal Streamflow Forecasting service[3] (SSF).

The AWO is a nationwide, grid-based service providing historical simulations and seasonal forecasts of key hydrological variables of the Australian surface water balance (Pickett-heaps & Vogel, 2022; Vogel et al., 2021). The service is underpinned by the Australian Water Resource and Assessment model (Frost, Andrew J, Shokri, 2021; Frost et al., 2018) and climate forcing is provided either by historical gridded climate data (Evans et al., 2020; Jones et al.,

---

[1] http://www.bom.gov.au/
[2] https://awo.bom.gov.au
[3] http://www.bom.gov.au/water/ssf/index.shtml

1    2009) or seasonal climate forecasts[4] (Wedd et al., 2022).

2    The SSF service (Feikema et al., 2018) provides seasonal forecasts of river discharge and natural inflows at specific

3    locations across Australia, coincident with river gauging stations and water storages respectively. A Bayesian Joint

4    Probability (BJP) statistical model (Zhao et al., 2016 and references therein) underpins the SSF service. BJP is

5    calibrated independently at each forecast location and is dependent on antecedent observed streamflow (a

6    representation of the catchment initial state) as the primary input data stream with which to forecast streamflow.

7    Antecedent observed streamflow is calculated from real-time hydrological observations. Data latency and data quality

8    problems of real-time observations are addressed through manual intervention and represents a significant barrier to

9    the full automation of the SSF service.

10    Replacing antecedent observed streamflow with AWO seasonal forecasts as the primary input data stream with which

11    to forecast seasonal streamflow will improve the flexibility and efficiency of forecast delivery for the SSF service.

12    This study investigates whether a gridded landscape hydrological model (AWRA-L), coupled with statistical post-

13    processing, generates acceptable seasonal streamflow forecasts relative to a locally calibrated BJP statistical model

14    based primarily on antecedent observed streamflow. A focus of this study whether dependency on real-time

15    hydrological observations as an input into the post-processing model can be removed without sacrificing forecast

16    performance. The catchment initial state has a significant impact on seasonal forecast skill due to hydrological

17    persistence (Hao et al., 2018; Van Dijk, Peña-Arancibia, et al., 2013). If AWRA-L provides an acceptable

18    representation of this initial state, inclusion of recent hydrological observations (an additional observed constraint)

19    becomes either optional or redundant. Historical hydrological observations are used to calibrate a statistical post-

20    processing model, providing a bias correction that is a function of the AWRA-L hydrological model hydrological

21    response only. A dependency on forecast post-processing in some form remains as aggregated runoff from AWRA-L

22    at the catchment scale is significantly biased when compared to observed river discharge (Yang & Wang, 2019, Bureau

23    internal report).

24    Many hydrological models targeting river discharge employ some form of hydrological post-processing to improve

25    simulation performance due to autocorrelated errors arising from hydrological persistence in river discharge (Bennett

26    et al., 2014; Lerat et al., 2020; Li et al., 2016; McInerney et al., 2017, 2020; Woldemeskel et al., 2018; T. Zhao et al.,

27    2015). BJP has also been used in the past as both a hydrological statistical post-processing model (Pokhrel et al., 2013)

28    and rainfall post-processing model (Ghajarnia et al., 2016; Schepen et al., 2018; P. Zhao et al., 2022a). Hydrological

29    post-processing models often rely on recent hydrological observations (antecedent observed streamflow) to provide

30    an error correction (Li et al., 2016; McInerney et al., 2020; Woldemeskel et al., 2018 and references therein).

31    This study provides an analysis of forecast performance across 449 forecast locations across Australia and within

32    different geographical regions and seasons of the water year (see Appendix B). Alignment by water-year is important

33    given the opposing seasonal pattern in rainfall/streamflow across northern and southern Australia. Rainfall in northern

34    Australia is dominated by the 'wet season', typical of tropical climates and occurs during the summer months of the

35    southern hemisphere (October-April). Rainfall in southern Australia is dominated by the temperate rainy season in the

---

[4] http://www.bom.gov.au/climate/outlooks/#/rainfall/summary

1    winter-spring months of the southern hemisphere (June-October).

2    **2.    Methodology**

3    The proposed modelling workflow to generate seasonal streamflow forecasts based on seasonal forecasts from a

4    gridded hydrological model is shown in Figure 1. The workflow includes existing operational systems that support

5    current seasonal forecasting services, as well as proposed downstream forecast post-processing system to generate

6    seasonal streamflow forecasts. A description of each modelling system and input datasets in Figure 1 is provided in

7    this section. A glossary of terms and variable definitions relevant to this paper are included in Appendix B.
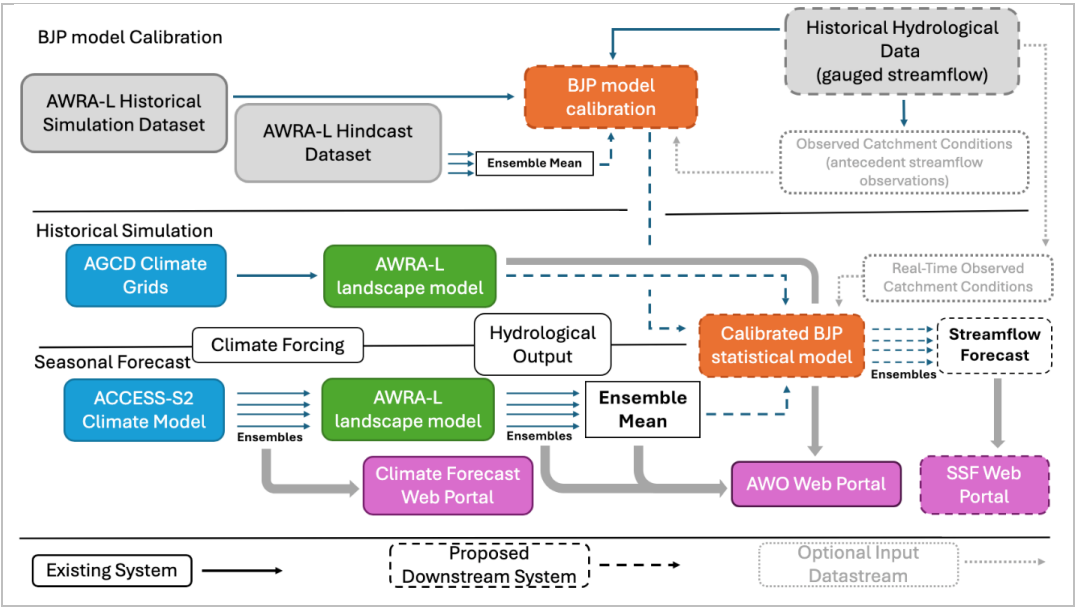


8    Figure 1: Modelling system workflow schematic incorporating seasonal climate forecasting, hydrological

9    forecasting and downstream BJP post-processing. Input datasets required for the statistical BJP model calibration are

10   also shown.

11   **2.1  Gridded Seasonal Climate and Hydrological Forecasts**

12   The Bureau maintains a nation-wide, grid-based hydrology monitoring and forecasting service of the Australian

13   surface water balance – The Australian Water Outlook (AWO). The AWO service consists of three separate

14   components: a historical simulation updated daily, seasonal hydrological forecasts and hydrological projections till

15   2100. This study is dependent on both historical simulations and seasonal forecasts provided by the AWO (Table 1).

| Service | Temporal resolution | Spatial Resolution | Update frequency | Forecast period | Lead-time interval | Ensemble based | Climate forcing | Reference |
|---------|---------------------|--------------------|------------------|-----------------|--------------------|----------------|-----------------|-----------|
| Historical Simulations | Daily | 5km2 | Daily | N/A | N/A | N/A | Gridded climate | Frost et al., 2018 |

| | | | | | | | (AGCD) | |
|---|---|---|---|---|---|---|---|---|
| Seasonal Forecasts | Monthly | 5km2 | Monthly | 3 months | 1 months | Yes | ACCESS-S2 seasonal climate forecasts | Pickett-heaps & Vogel, 2022 |

1 Table 1: Components of the Australian Water Outlook service on which seasonal forecasts of streamflow are
2 dependent.

3 The AWO seasonal forecasting service provides ensemble-based seasonal forecasts of key hydrological variables such
4 as soil-moisture, runoff, actual-evapotranspiration and potential evapotranspiration. The seasonal forecasts are
5 supported by two key modelling systems (Figure 1) - The Australian Water Resource and Assessment Landscape
6 Model (AWRA-L) and the Australian Community Climate Earth System Simulator – Seasonal, version S2 (ACCESS-
7 S2).

8

9 AWO is underpinned by AWRA-L (Frost et al., 2018) – a landscape model that represents the surface water balance
10 across Australia at a 5km spatial resolution and is run at a daily time-step. Key hydrological variables relevant to
11 seasonal streamflow forecasting are root-zone soil-moisture (combined soil layer of 0-1m thickness, herein referred
12 to as soil-moisture) and runoff. ACCESS-S (version S2) is a global multi-week to seasonal ensemble-based climate
13 model defined at a 60km2 resolution (Wedd et al., 2022) and is downscaled to a 5km2 spatial resolution (Griffiths et
14 al., 2023). ACCESS-S2 supports seasonal climate forecasting services provided by the Bureau and the necessary
15 climate input forcing for AWRA-L to generate seasonal hydrological forecasts available for the AWO. Further
16 information regarding the AWRA-L model and ACCESS-S2 model is provided in Appendix D.

17

18 **2.2 The Bayesian Joint Probability (BJP) statistical model for hydrological post-processing**
19 The Bayesian Joint Probability (BJP) model (and modelling package) is configured as either a statistical hydrological
20 model or a hydrological post-processor. The BJP model has been used as a statistical hydrological model for catchment
21 hydrology (Robertson & Wang, 2012; Wang et al., 2009; Wang & Robertson, 2011; T. Zhao et al., 2016) as well as a
22 rainfall post-processor (Schepen et al., 2018; Wang et al., 2019; P. Zhao et al., 2022b) and hydrology post-processor
23 (T. Zhao et al., 2015). The primary reason BJP was selected for this study is because BJP underpins the existing SSF
24 operational service, has been used previously as a hydrological post-processor and has advantages in relation to
25 operational forecasting (Appendix C).

26

27 The BJP model is defined by a multi-variate Gaussian distribution ($P(X)$) of a set of random variables ($X$). With fixed
28 (known) values assigned to a subset of random variables in $X$ (denoted as $X_1$), the conditional probability distribution
29 of the remaining random variables ($X_2$) can be evaluated. The BJP formulation and implementation used in this study
30 matches that of T. Zhao et al., 2016 which underpins the current SSF service and follows on from developments
31 described in Robertson & Wang, 2012; Wang et al., 2009; Wang & Robertson, 2011. Further details of the BJP
32 statistical model are included in Appendix C.

### 2.3 Forecasting streamflow from AWO seasonal forecasts

Two AWRA-L hydrological (input) variables were selected to generate seasonal streamflow forecasts – root-zone soil-moisture (herein referred to as soil-moisture) and runoff.

Runoff includes both surface runoff, baseflow and interflow and is considered most representative of relatively small, unimpaired catchments (Frost et al., 2018). Runoff does not include hillslope processes and impacts, lateral flow or routing. Seasonal forecasts of runoff indicate positive forecast skill at lead-time 0 for most of Australia, and positive skill for later lead-times for certain parts of the country and/or certain times of the year (Pickett-heaps & Vogel, 2022).

Root-zone soil-moisture is a representation of up-stream catchment conditions that impact streamflow (similar to that of observed antecedent streamflow). A soil-moisture forecast thus represents a future change in catchment conditions that will impact streamflow. Soil-moisture is included as an input hydrological variable as seasonal forecasts of soil-moisture exhibit greater forecast skill than runoff (Pickett-heaps & Vogel, 2022).

The generation of a single-dimension time-series from a gridded dataset follows the geospatial processing methodology as described in section 2.6. This geospatial calculation is applied to each AWO forecast (hindcast) ensemble member. The *mean* across all ensembles (Figure 1) is identified at each monthly lead-time.

NOTE: Forecast skill of runoff and soil-moisture (Pickett-heaps & Vogel, 2022; Vogel et al., 2021) is measured against an equivalent reference value obtained from an AWRA-L historical simulation with input forcing from historical climate grids (Table *3*). It does not relate to forecast skill arising from a comparison to streamflow.

### 2.4 Post-processing grid-based seasonal hydrological forecasts with BJP to generate seasonal streamflow forecasts

Figure 1 describes the modelling framework within which BJP fits as a down-stream forecasting post-processing system to generate seasonal streamflow forecasts. Figure 2 describes in more detail the application of BJP to post-process AWO seasonal forecasts for the generation of seasonal streamflow forecasts. A set of random variables $X$ (Appendix C) is partitioned as a set of predictor and predictand variables, $X_1$ and $X_2$ (Table 2).
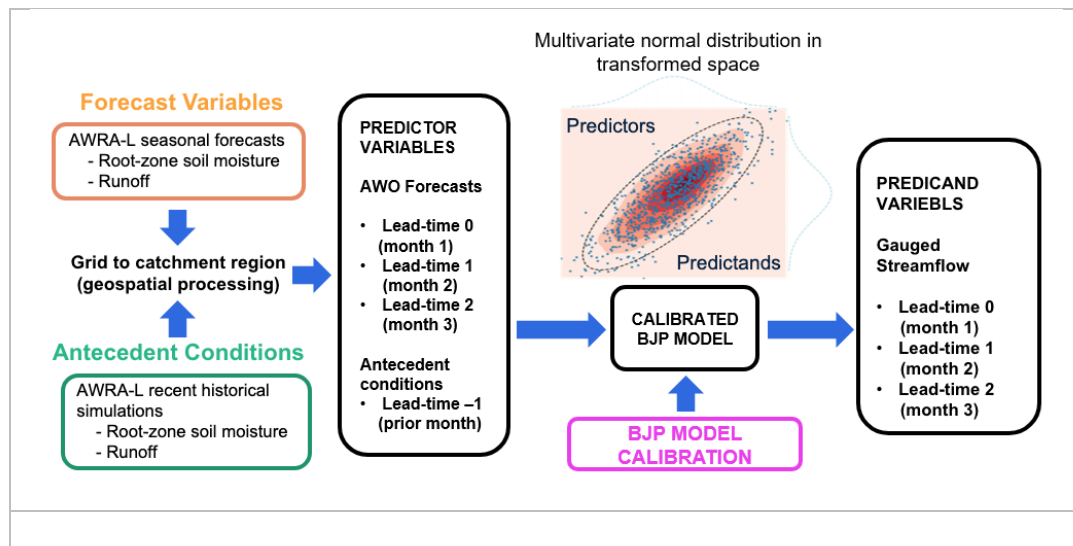
1    Figure *2*: Schematic diagram describing the statistical post-processing method to generate seasonal streamflow

2    forecasts from gridded seasonal hydrological forecasts from AWRA-L.

| Lead-Time | Predictor variables ($X_1$) | Predictand variables ($X_2$) |
|---|---|---|
| LT 0 | $F_{LT0}$ | $Q_{LT0}$ |
| LT 1 | $F_{LT1}$ | $Q_{LT1}$ |
| LT 2 | $F_{LT2}$ | $Q_{LT1}$ |
| LT -1 | $A_{LT-1}$ | N/A |

- Q (at lead-time LT) is observed streamflow (predictand variable)
- F (at lead-time LT) is an input forecast from AWO (predictor variable)
  - F may be assigned to either soil-moisture (sm) or runoff (qtot)
- A (at lead-time –1) is a variable representing antecedent conditions:
  - AWO historical simulation at lead-time –1 (either soil-moisture or runoff)
  - Observed antecedent streamflow (Qobs)

3         Table 2: Defined predictor and predictand variables of a BJP statistical model

4    The calibration of BJP requires the historical time-series of all random variables in *X* (Figure **1**, Figure **3**) and are

5    defined as follows:

6    - BJP predictor (*$X_1$*) variables:

7      - The AWO hindcast dataset. As described in section **2.7**, a hindcast dataset consists of a set of

8        retrospective forecasts generated within a defined period of time.

9      - The AWO historical simulation over the same hindcast period.

10   - BJP predictand (*$X_2$*) variables:

11     - Historical streamflow observations (see section **2.6**) over the same hindcast period
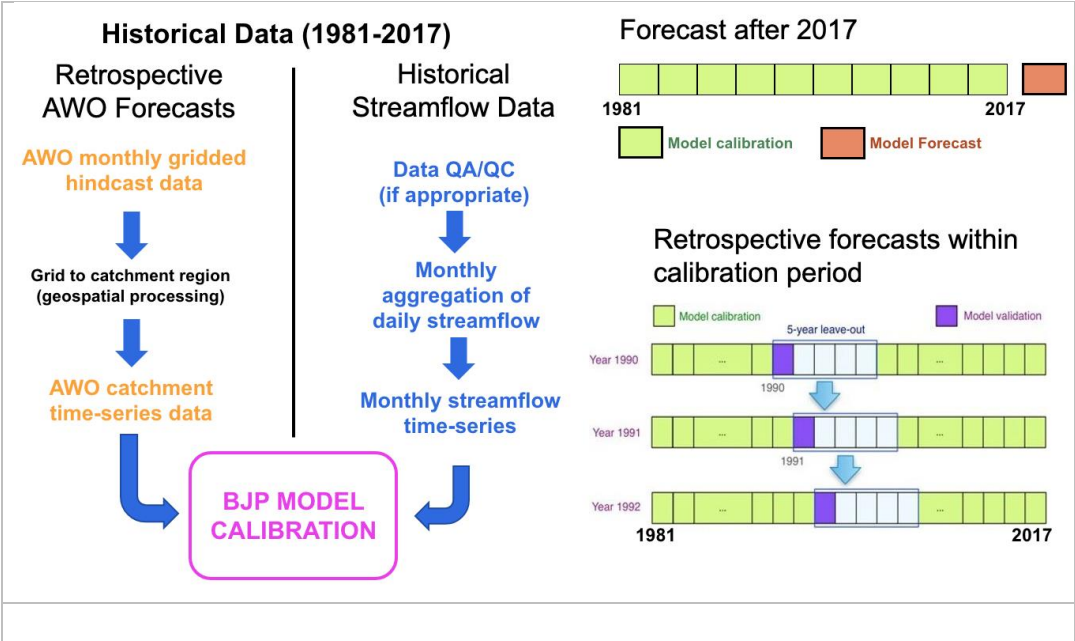
1    Figure *3*: Time-series datasets required for BJP model calibration. Application of a 5-year leave-out period when

2    generating retrospective forecasts (hindcasts).

3    AWRA-L predictor variable estimates are paired with the equivalent observed streamflow at each lead-time, with the

4    hindcast period (Table 3) dictating the data sample size available to calibrate BJP. Fixed values assigned to variables

5    in $X_1$ constrain the probable range of the unknown variables in $X_2$ (section 2.2) and are obtained by taking the *mean*

6    of the AWRA-L forecast ensembles at each lead-time. A single BJP model is defined for each site-month pair, as

7    recommended by T. Zhao et al., 2016. However, a single BJP model configured across all lead-times maintains the

8    strong temporal correlations in the conditional probability distributions of the predictand variables in $X_2$. This is

9    important if, for example, aggregated forecasts across successive months (e.g. 3-month totals) are required, as is the

10    case in the SSF service.

11    **2.5  Input datasets for the BJP model calibration**

12    This section describes the input datasets required for the BJP model calibration described in section 2.4.

13    **2.6  Hydrological data for forecast locations**

14    BJP model calibration is dependent on a historical record of hydrological observations. Forecast streamflow for

15    ungauged catchments was not attempted in this study.  Consequently, forecast locations are limited to:

16    •   Gauged catchments, for which volumetric streamflow observations are recorded

17    •   Major water storages (i.e. dams), for which historical estimates of natural inflows are available.

18    Figure 4 provides maps of forecast locations and associated up-stream catchments across Australia included in this

19    study. The selection of gauging stations for this study was based on the union between sites included in the SSF service
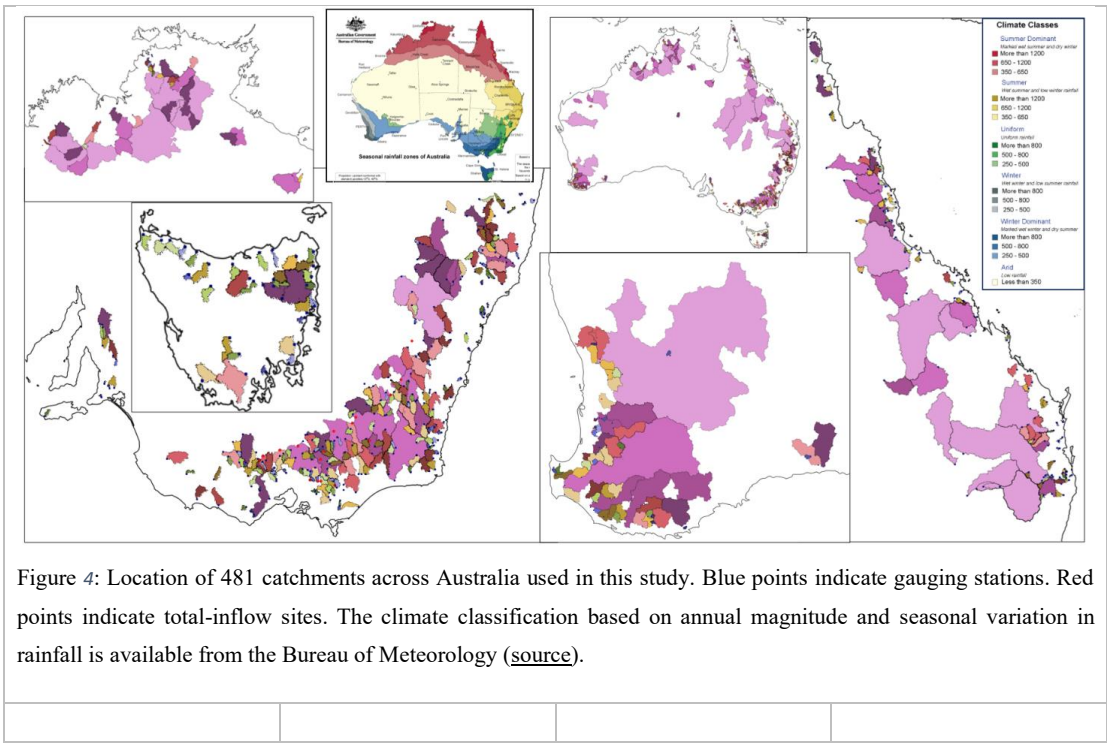
1    and sites included in the Hydrological reference Station (HRS) dataset (Amirthanathan et al., 2023). Implicit in this

2    selection of gauging stations are the following characteristics:

3    • Historical data time-series includes, as a minimum, the period: 1981-2017

4    • Hydrological data are of high-quality with adequate QA/QC having already been applied.

5    • Raw hydrological data typically originates as hourly data prior to temporal aggregation.

6    Natural inflows into 22 major water storages across Australia are provided by federal/state agencies. In this

7    manuscript, forecast location refers to either a gauging river station or a major water storage.



Figure *4*: Location of 481 catchments across Australia used in this study. Blue points indicate gauging stations. Red points indicate total-inflow sites. The climate classification based on annual magnitude and seasonal variation in rainfall is available from the Bureau of Meteorology (source).

8

9    Time-series data for all predictor variables based on AWRA-L (runoff and soil-moisture, Figure 3) is required for

10   each upstream catchment in Figure 4. Extraction of gridded data requires a shapefile that delineates the upstream

11   catchment of each forecasting location. A single-dimensional data time-series is calculated either from a spatial

12   weighted sum (flux variable such as runoff) or weighted mean (state-variable such as soil-moisture). The resulting

13   time-series is paired with concurrent hydrological observations from a forecasting location (Figure 3).

14   **2.7 Gridded datasets – AWRA-L hindcast dataset and historical simulation for BJP model calibration**

15   A hindcast dataset consists of retrospective forecasts (or hindcasts) generated within a defined period and paired with

16   a corresponding reference dataset (e.g. hydrological observations). In this study, an AWRA-L hindcast dataset is

17   required for the calibration of the BJP models, conditioned against the paired streamflow observations.

1

2  The AWRA-L hindcast dataset (Table 3) underpins the forecast verification provided by the AWO (Pickett-heaps &

3  Vogel, 2022). The dataset is derived from the ACCESS-S2 calibrated (downscaled) hindcast dataset, the size of which

4  (number of hindcasts) is constrained primarily by available computational resources. Both hindcast datasets are key

5  to    the    assessment    of    forecast    skill    of    AWRA-L    and    ACCESS-S2    respectively.

6

| | Time period (years) | Number of hindcasts | Temporal resolution | Temporal duration | Ensembles | Reference dataset |
|---|---|---|---|---|---|---|
| AWO | 1981-2017 (37) | 444 | 1-month | 6-months | 27 | AWRA-L historical simulations (Frost et al., 2018) |

7  Table 3: Hindcast specifications for AWO and ACCESS-S2

8  The AWRA-L historical simulation is also used to define an input predictor variable, as indicated in Figure 2.

### 2.8  The streamflow hindcast dataset and streamflow forecast verification

10  Streamflow forecast verification requires a suitable hindcast dataset. Generating retrospective forecasts is the same as

11  that for real-time forecasts (section 2.4), except for a 5-year leave-out period (Figure 3) that ensures streamflow

12  observations used for validation remain independent of a BJP calibration (Woldemeskel et al., 2018; T. Zhao et al.,

13  2016). A 5-year leave-out period removes the impact of long-term hydrological persistence in baseflow that could

14  violate this assumption of independence. A leave-out period for each hindcast date implies a separate calibration

15  dataset and calibration run. Significant computational resources are necessary to generate a hindcast dataset, consisting

16  of 444 retrospective forecasts, for 481 forecasting locations.

17  Forecast verification is achieved through the calculation and analysis of verification metrics (Lerat et al., 2020; Pickett-

18  heaps & Vogel, 2022; Vogel et al., 2021; Woldemeskel et al., 2018; T. Zhao et al., 2016). Two primary forecast

19  metrics were used in this study - The Cumulative-Rank-Probability-Score (CRPS) and forecast reliability. Further

20  information regarding these forecast metrics is included in Appendix E.

21  CRPS estimates are expressed in units of river discharge (ML or mm). To improve interpretability, the Cumulative-

22  Rank-Probability-Skill-Score (CRPS-S) compares a CRPS estimate from a forecast model to a reference model (e.g.

23  a historical climatology or a defined benchmark model) as a ratio (Woldemeskel et al., 2018). The subsequent

24  assessment of improved forecast skill compared to a reference model is straight forward:

25  • CRPS-S > 0 indicates an improvement in forecast skill.

26  • CRPS ~ 0 indicates no improvement in forecast skill.

27  • CRPS < 0 indicates forecast skill deterioration.

28  The historical climatology of streamflow used in this study is the same as the hindcast period: 1981-2017.

29  Seasonal variation in forecast skill across Australia is analysed with respect to the local water year rather than the

30  calendar year, given opposing seasonal variation in streamflow between northern and southern Australia. The local

1   water-year (Appendix B) is defined by the minimum mean streamflow based on a three-month moving window across

2   all 12 months of the calendar year. The central month in which this minimum three-month flow occurs is set to *month*

3   *2* of the 12-month water year. Monthly forecast skill scores (relative to the calendar year) are then adjusted accordingly

4   for each site, resulting in the correct alignment of forecast skill relative to the defined water year at each site. Categories

5   were defined to identify different periods of the water-year (Table 4).

| Defined period in the water-year | Months of the water year |
|---|---|
| Low-flow season | [1,2,3] |
| Rising-limb season | [4,5,6] |
| High-flow season | [7,8,9] |
| Falling-limb season | [10,11,12] |
|  |  |
| Low flow period (6 months) | [1,2,3,4,5,12] |
| High flow period (6 months) | [6,7,8,9,10,11] |

6   Table 4: Defined periods of the water-year.

7   The results in section 3 detail an analysis of the median forecast skill and median difference in forecast skill across a

8   sample of sites. The median difference is calculated from the distribution of pairwise differences (Table *5*) between

9   two samples of sites (not the difference between two median estimates). Confidence intervals of the median can

10  determine if the median (or median difference) is statistically different from zero. Bootstrap sampling of the median

11  estimate of a sample was applied to generate confidence intervals.

| Sample A (n sites) | Sample B (n sites) | Pairwise differences (n sites) |
|---|---|---|
| $\xi_{A1}$ | $\xi_{B1}$ | $d_1 = \xi_{A1} - \xi_{B1}$ |
| $\xi_{A2}$ | $\xi_{B2}$ | $d_2 = \xi_{A1} - \xi_{B1}$ |
| $\xi_{A3}$ | $\xi B_3$ | $d_3 = \xi_{A1} - \xi_{B1}$ |
| --- | --- | --- |
| $\xi_{An}$ | $\xi_{Bn}$ | $d_n = \xi_{A1} - \xi_{B1}$ |
| Median skill: $\xi_{m,A}$ | Median skill: $\xi_{m,B}$ | Median difference: $d_{m\xi}$ |

12  Table 5: Median forecast skill estimates from a sample of sites from two model,

13  A & B. The variable ξ represents a skill metric (e.g. CRPS-S). Samples consist of

14  the same number of sites.

15  **2.9  BJP model configuration experimentation**

16  As described in section 2.4, different AWRA-L forecast variables can be assigned as BJP predictor variables at all

17  forecast lead-times ($F_{LTX}$, Table 2) and the lead-time representing antecedent conditions:

| Forecast lead-times | Antecedent catchment conditions |
|---|---|
| • Runoff (qtot) | • Runoff (qtot) |

- Root-zone soil-moisture (sm)
- Root-zone soil-moisture
- Observed streamflow (Qobs)

1

2 The configurations of predictor variables are summarized in Table 6:

| AWRA-L only | AWRA-L + observed streamflow | Benchmark model |
|---|---|---|
| $Q(t) = BJP(A_F(\alpha, t))$ <br> $Q(t) = BJP(A_F(\alpha, t), A_H(\alpha, t-1))$ | $Q(t) = BJP(A_F(\alpha, t), Qobs(t-1))$ | $Q(t) = BJP(Qobs(t-1))$ |
| <ul><li>$t$ – Forecast initialisation time<ul><li>$L$ – Defined forecast lead-times</li></ul></li><li>$Q(t)$ – Forecast streamflow at time $t$ across lead-times defined by $L$</li><li>$BJP(\theta)$ – Calibrated BJP model with input predictor variables $\theta$</li><li>$\theta$ predictor variables:<ul><li>$A_F(\alpha,t)$ – AWRA-L forecasts at time $t$ across lead-times defined by $L$</li><li>$A_H(\alpha,t-1)$ – AWRA-L historical simulation at time $t-1$</li><li>$Qobs(t-1)$ – Antecedent observed streamflow at time $t-1$</li></ul></li><li>$\alpha$ AWRA-L variables:<ul><li>Root-zone soil-moisture</li><li>Runoff</li></ul></li></ul> | | |

3 Table 6: Summary of the BJP model configurations used in this study.

4 Separate BJP model configurations were defined for a series of modelling experiments, with the objective to identify
5 a configuration that maximized forecast skill. Table 7 listed the 8 different model configurations used in this study.

| | Lead-time | 0 | 1 | 2 | -1 |
|---|---|---|---|---|---|
| Root-zone soil-moisture | Config-01 | sm | sm | sm | --- |
| | Config-02 | sm | sm | sm | sm |
| | Config-03 | sm | sm | sm | Qobs |
| Runoff | Config-04 | qtot | qtot | qtot | --- |
| | Config-05 | qtot | qtot | qtot | qtot |
| | Config-06 | qtot | qtot | qtot | Qobs |
| Hybrid | Config-07 | qtot | qtot | qtot | sm |
| Benchmark Model | Config-08 | --- | --- | --- | Qobs |

6 Table 7: BJP model configurations. Root-zone soil-moisture is denoted as 'sm'. Runoff is
7 denoted as 'qtot'. Antecedent observed streamflow is denoted as 'Qobs'.

8 For each model configuration in Table 7 and subsequent model experimentation, a full verification analysis was
9 performed across 481 sites. The benchmark model in Table 7 represents a similar model configuration to that used in
10 the current SSF operational service. Soil-moisture and runoff are not included together in a single BJP model due to
11 a significant increase in the number of parameters in the BJP model that must be optimized.

### 2.10 Selection of the best-performing BJP model configuration

Different predictor variable configurations (Table 7) will generate variation in forecast performance as measured by CRPS-S, where either runoff or soil-moisture are likely to provide superior forecast performance on a consistent basis at individual sites during different times of the water year. Identifying the configuration that maximizes forecast performance at each site-month pair will improve forecast performance overall and provide a unique predictor configuration to be used in operational forecasting.

Identification of superior forecast performance is ostensibly achieved by comparing forecast skill as measured by CRPS-S. However, this comparison itself should be subject to independent verification. For a unique hindcast ($\mathbf{d}$) drawn from the 1981-2017 hindcast period, the calculated CRPS-S values for all predictor variable configurations should be independent of this hindcast $\mathbf{d}$. This ensures that the choice of optimal configuration for hindcast $\mathbf{d}$ is independent of the CRPS-S calculation and subsequent comparison. In a similar fashion to the 5-year leave-out period described in Figure 3, identification of the optimal configuration must be completed for each hindcast $\mathbf{d}$ in the hindcast dataset for each site-month pair.

The following methodology is used to identify the optimal predictor variable configuration for each site-month pair and to calculate the forecast skill resulting from the selection of an optimal predictor:

- A new configuration is defined for which a new hindcast dataset ($\mathbf{H}$) is required for each site-month pair. A label (e.g. 'MAXSCR') is assigned to this new configuration.
- The hindcast dataset $\mathbf{H}$ for MAXSCXR consists of individual hindcasts ($\mathbf{h}$) selected from existing hindcast datasets across a selection of existing configurations ($\mathbf{Y}$) in Table 7.
- For each hindcast date $\mathbf{d}$, a 5-year leave-out period is defined (Figure 3).
- Individual hindcasts within this period are removed and a CRPS-S is calculated from the remaining hindcasts for each selected BJP predictor variable configuration in $\mathbf{Y}$.
- The configuration with the highest CRPS-S value across the configurations in $\mathbf{Y}$ dictates the specific hindcast $\mathbf{h}$ selected for the individual hindcast date $\mathbf{d}$. This hindcast $\mathbf{h}$ is added to the new hindcast dataset ($\mathbf{H}$) for MAXSCR.
- This process is repeated for every hindcast date $\mathbf{d}$ in the hindcast period so that a full hindcast dataset for MAXSCR is compiled as a combination of hindcasts from the original model configurations defined in $\mathbf{Y}$ (Table 7)
- The standard forecast performance evaluation is then applied to the new hindcast dataset $\mathbf{H}$.

The best-performing predictor configuration for each hindcast date $\mathbf{d}$ is unlikely to be the same for all hindcast dates. Thus, the *proportion of hindcasts* favoring each predictor configuration is calculated. The proportions are combined for configurations based either on runoff or soil-moisture variables. The 'hybrid' configuration (config-07) was excluded as there was no apparent benefit from including this additional configuration. The confidence interval for this proportion will identify whether the proportion is significantly different from 0.5 (no preference). Confidence intervals were calculated using the Python package 'statsmodels' (Seabold & Perktold, 2010), following the advice

1     from (Brown et al., 2001) where the Wilson Score Interval for proportions is recommended.

2     **3. Results**

3     Results presented in this section aim to determine:

4     •   Whether seasonal streamflow forecasts from a gridded hydrological model alone combined with statistical
5        post-processing provides acceptable forecast skill relative to a model based on antecedent streamflow
6        observations only (benchmark model).

7     •   Whether observed antecedent streamflow remains necessary to maintain forecast skill.

8     •   Whether AWRA-L provides additional forecast skill relative to the benchmark model from improved
9        representation of initial catchment conditions and climate forcing from ACCESS-S2.

10     The analysis of forecast skill focuses on:

11     •   Median forecast skill derived from a sample of forecast locations and the (significant) difference in forecast
12        skill estimates between two samples (Table 5). Differences in forecast skill is calculated from the pairwise
13        differences in skill at individual sites within two samples.

14     •   The distribution of forecast skill from a sample of forecast locations, represented as empirical cumulative
15        distribution function (CDF) curves (Appendix A). Distributions in forecast skill are stratified across four
16        seasons of the water-year (defined in Table 4) and defined spatial regions (Table *8*).

17     **3.1 Success in BJP calibration runs**

18     Out of 481 forecast locations (section 2.5), BJP calibration runs failed for between 20-30 sites (due to a single or

19     multiple individual BJP calibration failures). BJP calibration failure is not unexpected and occurs in BJP calibration

20     runs for the SSF operational service, usually due to a failure to converge to an optimized parameter solution of the

21     BJP model for a single calibration run. Efforts to resolve these failures were deemed unwarranted due to the low

22     percentage of sites with calibration failures. The objective of this study is to establish the potential of using gridded

23     hydrological forecasts to generate seasonal streamflow forecasts across Australia as described in Figure 1. Options to

24     either overcome or mitigate calibration failures will be addressed in the development of an operational system.

25     A full hindcast dataset for between 450-460 sites was generated, depending on the BJP model configuration used

26     (section 0). The intersection across successful calibrations for all configurations resulted in 449 forecast locations.

27     **3.2 Forecast skill across all BJP configurations**

28     Figure 5 provides the median forecast skill of each BJP model configuration in Table 7 (section 2.9) for each month

29     of the water-year. The CRPS-S color-bar (Figure 5) is centered on the annual-average forecast skill of the benchmark

30     model. Also included is the *median difference in forecast skill* calculated from the pairwise differences in forecast

31     skill (Table *5*) relative to the benchmark model. Finally, a binary panel plot is included indicating whether there is a

32     statistically significant shift forecast median skill (section 2.8). Appendix A*1* provides the CDF plots ('AUS' region)

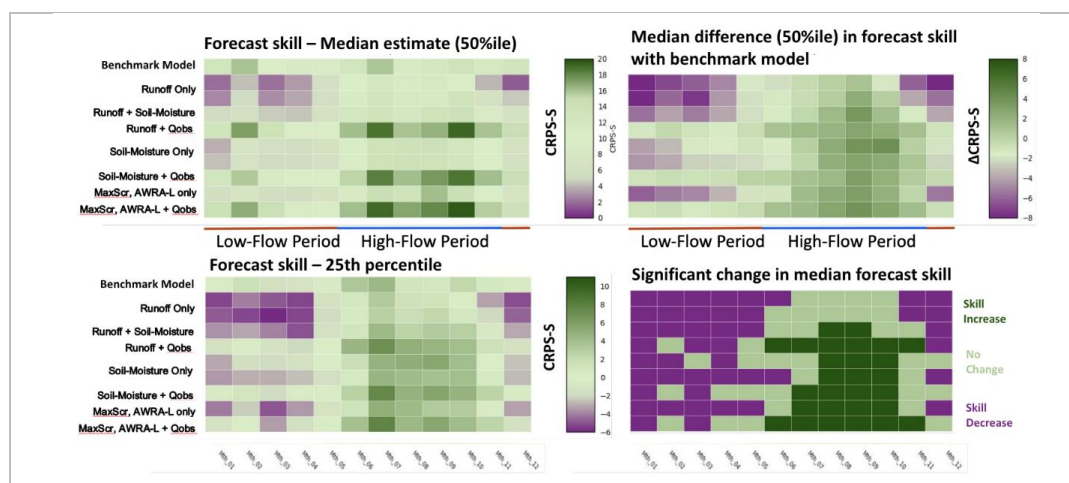33     for a selection of model configurations.

1   Figure *5*: Median forecast skill (CRPS-S) for all BJP model configurations (Table 7) at lead-time 0. Median difference

2   in forecast skill (CRPS-S) with the benchmark model. Forecast skill of the 25 %ile. Significant shift in median forecast

3   skill relative to the benchmark model. 'Runoff Only' refers to the two BJP model configurations that include runoff

4   only. 'Soil-Moisture Only' refers to the two BJP model configurations that include root-zone soil-moisture only.

5   Features apparent in Figure 5 at lead-time 0 are summarised as follows:

6   *BJP configurations with AWRA-L forecast variables only:*

7   • There is a clear reduction in forecast skill during the low-flow months of the year (statistically significant at
8     the 5% level).

9   • There is an increase in forecast skill for the high-flow season (months 7-9). This increase is statistically
10    significant for soil-moisture.

11  • Similar median forecast skill exists between configurations that include or exclude an AWRA-L predictor
12    variable at lead-time –1 (antecedent conditions). An exception is the case when soil-moisture at lead-time –
13    1 is combined with runoff for the forecast lead-times.

14  • Soil-moisture generates an increase in median forecast skill in the high-flow months whereas runoff does
15    not.

16  • Forecast skill at the 25th percentile for soil-moisture is clearly higher than that for runoff during the low-flow
17    months.

18  • The proportion of sites with positive forecast skill is similar for three out of four seasons of the water-year
19    (Appendix A1) when compared to the benchmark model. The low-flow season is an exception.

20  *BJP configurations with AWRA-L forecast variables and antecedent observed streamflow:*

21  • Forecast skill during the low-flow months of the year is similar to that of the benchmark model.

22  • Forecast skill during the high-flow period increases compared to the benchmark model.

23  • Across Australia (AUS region), the CDF curves of CRPS-S from AWRA-L combined with antecedent

1    observations closely match the CDF curve of the benchmark model (Appendix A2).

2    Analysis of forecast performance at individual site-month pairs revealed variation in CRPS-S depending on whether

3    soil-moisture or runoff is used as a predictor variable. The AWRA-L input variable that maximises forecast skill for

4    each site-month pair was identified following the method in section 2.10. Two new configurations were defined:

5    • MAXSCR – Maximum forecast skill across the four BJP model configurations using AWRA-L predictor

6       variables only (config-01, config-02, config-04, config-05). The 'hybrid' configuration (config-07) was

7       excluded (Table 7) as no apparent benefit was identified by including this additional configuration.

8    • MAXSCR-Q – Maximum forecast skill across the two BJP model configurations using AWRA-L predictor

9       variables and antecedent observations (config-03, config-06, Table 7).

10   The median forecast skill for MAXSCR and MAXSCR-Q are included in Figure 5. Results for MAXSCR and

11   MAXSXR-Q are summarised as follows:

12   • While remaining below the benchmark model, median forecast skill of MAXSCR improves during the low-

13      flow months of the water year.

14   • All six months of the high-flow period (water-year) indicate either no change or a significant increase in

15      median forecast skill of MAXSCR when compared to the benchmark model.

16   • Inclusion of antecedent observed streamflow (Qobs) results in a significant positive shift in median forecast

17      skill (MAXSCR-Q) in the high-flow season and similar forecast skill to the benchmark model in the low-

18      flow period.

19   These results indicate that inclusion of antecedent observed streamflow, while beneficial, is not essential to achieve

20   acceptable forecast skill. Only for applications focused on low flows might inclusion of antecedent observations be

21   warranted at individual forecast locations. At other times of the year, factors unrelated to forecast performance may

22   influence the decision to include antecedent observations (section 4.4).

23   The CDF curves of CRPS-S for MAXSCR are included in Appendix A3 and Appendix A5 ('AUS' region). There

24   remains reduced forecast skill in the low-flow season for MAXSCR when compared to the benchmark model. The

25   selection of the best-performing AWRA-L variable is insufficient to overcome limitations in forecast skill from

26   AWRA-L during the low-flow season. However, MAXSCR closely matches the CDF curve of the benchmark model

27   in the high-flow season. The CDF curve of MAXSCR-Q is included in Appendix A5 ('AUS' region), where it closely

28   matches benchmark model CDF curve in the low-flow season. In the high-flow season, the positive shift in the CDF

29   curve for MAXSCR-Q represents an increase in forecast skill relative to the benchmark model, attributable to forcing

30   from ACCESS-S2 given initial catchment conditions at the catchment scale are accounted for.

31   3.3  Forecast skill across different regions in Australia – Lead-time 0
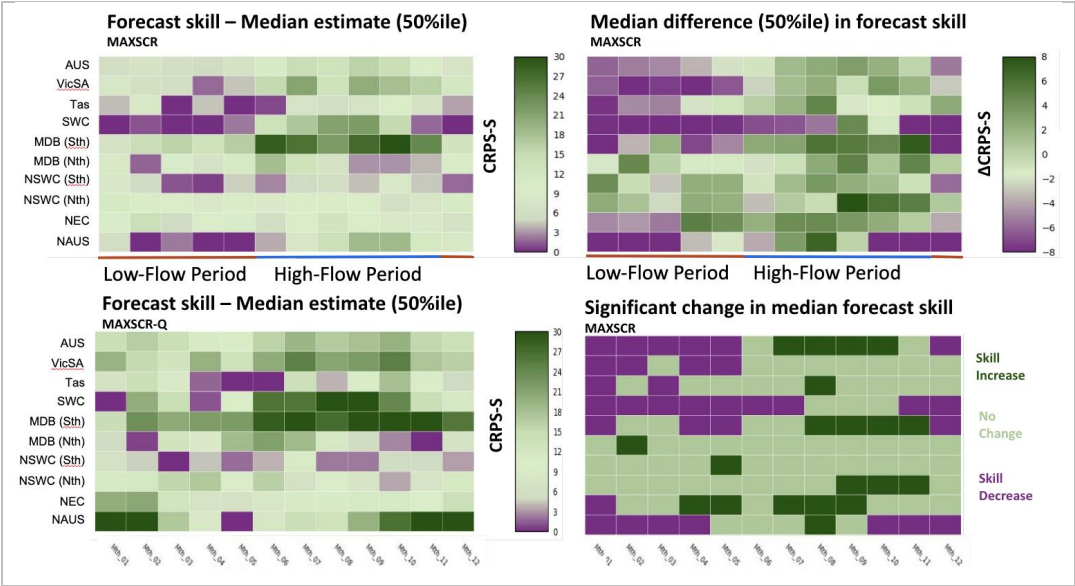
1    Figure 6: Median forecast skill (CRPS-S) for each defined spatial region (Table 8) for 'MaxScr' and 'MaxScr-Q' at

2    lead-time 0. Median difference in forecast skill between MaxScr and the benchmark model. Significant shift in median

3    forecast skill for MaxScr relative to the benchmark model.

| Region ID | No sites | Region Name | Original Geofabric Drainage Division(s) | Dominant seasonal rainfall characteristics |
|---|---|---|---|---|
| Aus | 449 | Australia | N/A | N/A |
| Vic-SA | 56 | Victoria/South Australia | South-East Coast (Victoria) & South Australian Gulf | Winter dominated |
| Tas | 28 | Tasmania | Tasmania | Winter dominated |
| SWC | 55 | South-West Coast (WA) | South-West Coast | Winter dominated |
| MDB (Sth) | 90 | Southern MDB | Murray-Darling Basin (south of 32S) | Winter dominated |
| MDB (Nth) | 39 | Northern MDB | Murray-Darling Basin (north of 32S) | Transitional zone |
| NSWC (Sth) | 27 | New South Wales Coast (south) | South-East Coast (NSW) (south of 32S) | Not well defined |
| NSWC (Nth) | 45 | New South Wales Coast (north) | South-East Coast (NSW (north of 32S) | Transitional zone |
| NEC | 66 | North-East Coast (QLD) | North-East Coast (Queensland | Summer dominated |
| NAUS | 43 | Northern Australia | Regions included: Tanami-Timor Sea | Summer dominated in |

| | | (includes central Australia) | Coast, Carpentaria Coast, Pilbara-Gascoyne, North-Western Plateau, Lake Eyre Basin<br>* No sites located in the South-West Plateau drainage division | northern Australia. Erratic in central Australia. |
|---|---|---|---|---|

1   Table 8: Defined spatial regions across Australia, based on the Geofabric (v3.4) drainage divisions.

2   Figure 6 provides a summary of median forecast skill across the water year. Median forecast skill for 'MAXSCR'
3   and 'MAXSCR-Q' is plotted for different regions across Australia. Regions are roughly ordered from south ('winter
4   dominated rainfall') to north ('summer dominated rainfall'). Forecast regions were derived from the Geofabric
5   Drainage Divisions (v3.4) and modified depending on the number of forecast locations within each division. The
6   Murray-Darling Basin drainage division is split into a northern and southern section, coincident with 32S and the
7   Lachlan basin. Similarly, the South-East Coast (NSW) drainage division is split into two at 32S. The NAUS region
8   represents an amalgamation of drainage divisions in northern, north-western and central Australia (Table 8). Most
9   forecast locations therein experience the distinct wet and dry seasonal variation of northern Australia.

10   Variation in median forecast skill across different regions (Table 8) is summarised as follows:

11   *MAXSCR: Improvements in forecast skill*

12   • During the high-flow period, locations in the Southern Murray-Darling Basin exhibit the highest forecast
13       skill. Performance is higher relative to Victoria (south of the Great Dividing Range), the east-coast of NSW
14       and the northern Murra-Darling Basin.
15   • Forecast skill in Tasmania is reduced compared to other regions in south-east Australia, as is the case for the
16       benchmark model.
17   • Forecast skill is reasonable along the entire eastern coast of Australia (east of the Great Dividing Range: –36
18       to –10 degrees latitude).
19   • Forecast skill is acceptable during the wet-season (high-flow period) across northern Australia.

20   *MAXSCR: Reductions in forecast skill*

21   • Reductions in forecast skill occur primarily in the low-flow period, with two regions showing particularly
22       large reductions relative to the benchmark model – The South-West Coast of Western Australia (SWC) and
23       the dry-season of northern Australia (NAUS).
24   • The Vic-SA spatial region also exhibits a reduction in forecast skill relative to the benchmark model, despite
25       lying adjacent to the southern Murray-Darling Basin.
26   • The North-East coast of Queensland (NEC) does not show a reduction in forecast skill during the 'low-flow'
27       period (northern Australia dry season).

28   MAXSCR-Q:

29   The inclusion of observed antecedent observations has the following impact:

1 • Forecast skill is enhanced for the southern Murray-Darling Basin, the Vic-SA region, and the high-flow
2 season (winter) of SWC during the high-flow period.
3 • Forecast skill appears somewhat unchanged for Tasmania, NEC and the northern Murray-Darling Basin.
4 • Forecast skill is maintained or enhanced during the low-flow season of SWC and across Northern Australia
5 during the dry-season (excluding NEC).

6 The CDF curves of CRPS-S for a selection of BJP model configurations (including and excluding Qobs) for the spatial

7 regions defined in Table *8* are included in Appendix A*1*, Appendix A*2*, Appendix A*3* and Appendix A*5*. The results

8 mirror those of Figure 6, where reductions in forecast skill relative to the benchmark model primarily occur in the

9 regions SWC, NAUS and Vic-SA. Apart from these regions, the proportion of sites with positive forecast skill in the

10 low-flow season is similar to the benchmark model. For all seasons of the water-year, SWC region is the most

11 challenging region to forecast given skill is reduced relative to the benchmark model. As would be expected, inclusion

12 of Qobs in the BJP model configuration helps overcome these challenges.

13 **3.4 Forecast skill at all lead-times**

14 Figure 7 provides a summary of median forecast skill for all lead-times and spatial regions (Table *8*), this time plotted

15 as forecast skill categories for comparative purposes. Model configurations MAXSCR, MAXSCR-Q and the

16 benchmark model are included. For each lead-time, the forecast month along the x-axis is interpreted as the month (of

17 the water-year) when that month is forecast at the specified lead-time. The forecast issue month is determined by

18 subtracting the number of lead-times. For lead-time 0, the forecast issue month is identical to the forecast month.
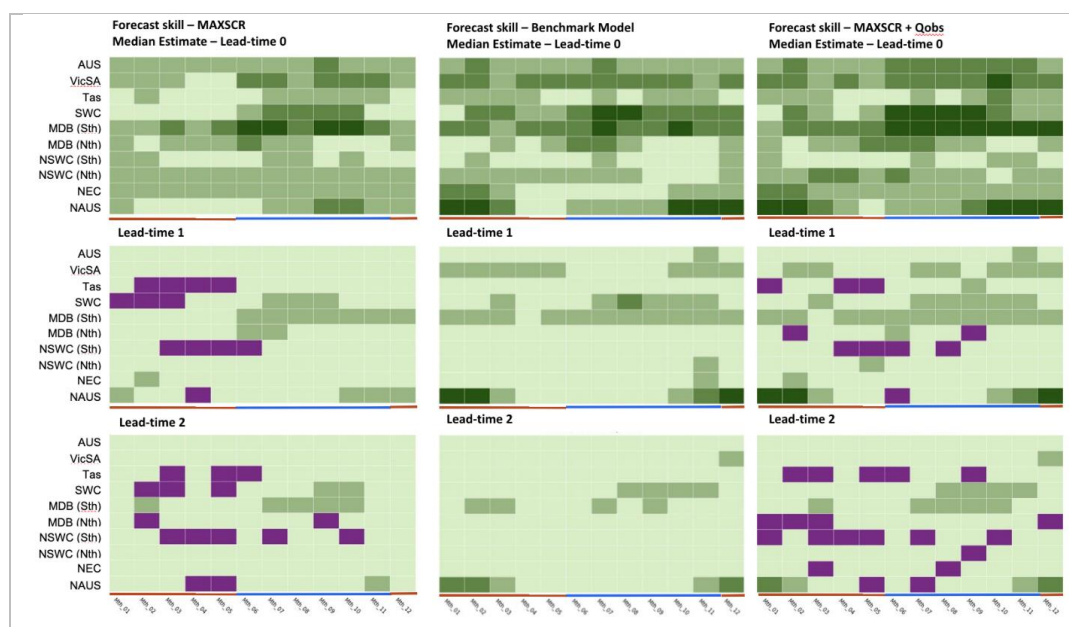
| Positive forecast skill: | No forecast skill: |
|---|---|
| **CRPS-S >25%** | CRPS-S: [-5%-5%] |
| **CRPS-S: [15%-25%]** | Negative forecast skill: |
| CRPS-S: [5%-15%] | **CRPS-S < -5%** |

Figure *7*: Forecast skill-scores (CRPS-S) at all monthly lead-times (0,1,2) for different BJP model configurations and different geographical regions. Forecast skill categories are shown to aid in the comparison in forecast skill with lead-time.

The fall in forecast skill for lead-times 1 and 2 relative to lead-time 0 is entirely expected (Figure 7). Other notable features include:

- Positive forecast skill for the southern Murra-Darling Basin is maintained through the high-flow period in lead-time 1 and the high-flow season in lead-time 2.
- Positive forecast skill is maintained during the dry season of Northern Australia (NAUS) with the inclusion of antecedent observations. This is reflective of strong persistence in streamflow during this time of the year. The inter-annual variability in this persistence is absent in AWRA-L.

The CDF curves for CRPS-S for MAXSCR and MAXSCR-Q are included in Appendix A*5*, Appendix A*6* and Appendix A*7*. CDF curves for MACSCR across all three lead-times for comparative purposes are included in Appendix A*8*. Apart from regions exhibiting very strong hydrological persistence in the low-flow season (NAUS, SWC), inclusion of Qobs as a predictor variable has little benefit relative to the benchmark model at lead-times 1 and 2. Most regions indicate a similar proportion of sites with positive skill for MAXSCR/MAXSCR-Q relative to the benchmark model.

### 3.5 Frequency in which AWRA-L predictor variable configurations maximize forecast skill

Section 2.10 describes the method to identify the AWRA-L predictor variable that maximises forecast skill, where a proportion of individual hindcasts for which soil-moisture and runoff maximises CRPS-S is calculated for each site-month pair. Figure 8 provides the median estimate of this proportion for the sample of sites in each region as well as the proportion of sites for which no significant preference exists (i.e. at a given site-month pair, an equal proportion of hindcasts favor either runoff or soil-moisture). Results shown are for MAXSCR.

Relatively few forecast locations (<~20%) in any given region indicate equal forecast performance between soil-moisture and runoff across the water-year. For most site-month pairs, either soil-moisture or runoff consistently maximises forecast skill. For certain months of the water-year across most regions, an equal proportion of sites favor either soil-moisture or runoff. The low-flow period indicates a preference for soil-moisture whereas relatively few regions and/or months of the water-year favor runoff.

Figure 8 also presents the median forecast skill for the sample of sites that identify either soil-moisture or runoff as maximising forecast skill. Overall, a greater proportion of sites indicate soil-moisture as providing increased forecast

1 skill relative to runoff. However, for the remaining sites that preference runoff, a clear increase in forecast skill is

2 evident. This is particularly evident in the southern Murray-Darling Basin, Vic-SA region, SWC and northern

3 Australia in the high-flow season. A few regions in certain months of the water-year are to be treated with caution due

4 to a small sample size (e.g Tasmania, MDBNth and NSWCsth).

5 The CDF curves of CRPS-S for the sample of sites favoring either runoff or soil-moisture are included in Appendix

6 A4. A clear difference in the distribution of CRPS-S is apparent in the high-flow season across Australia as well as

7 for the regions Vic-SA and Tas. For MDBSth, there is a noticeable difference in three out of four seasons of the water-

8 year (excluding the low-flow season). Surprisingly, there is also a difference in the CRPS-S distributions in NAUS

9 during the low-flow season and falling-limb season.
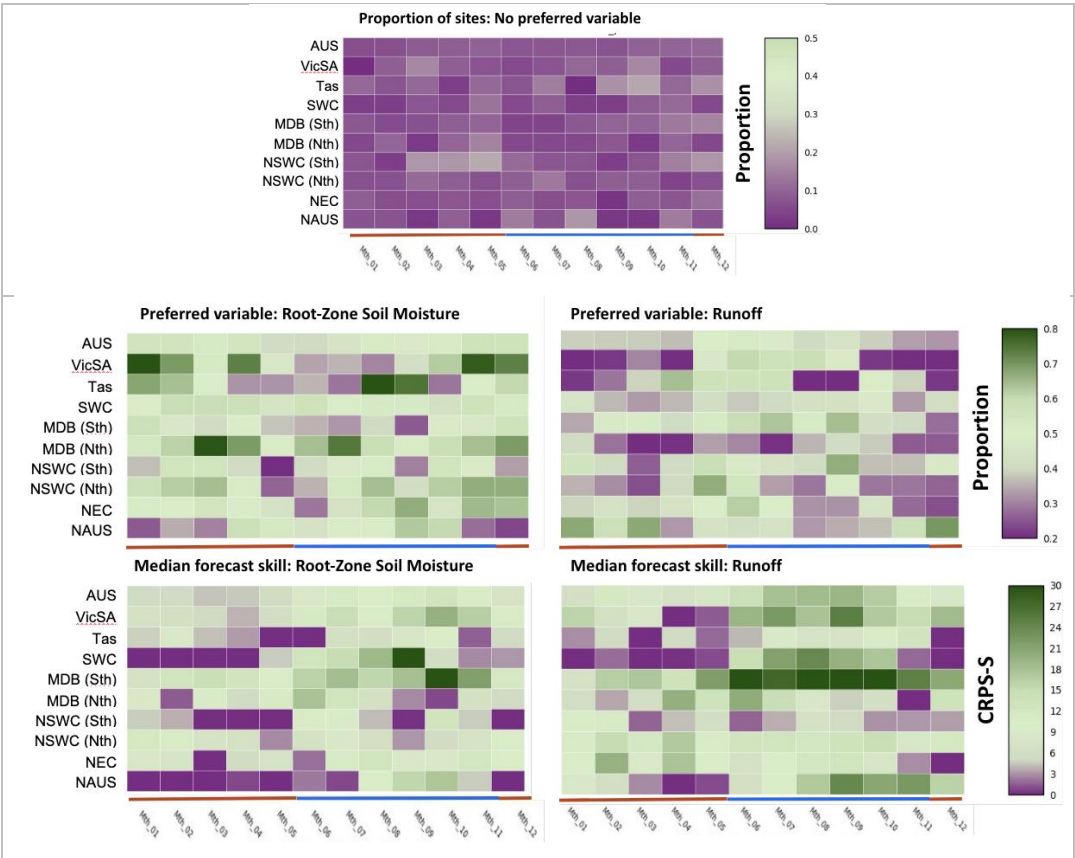


10 Figure *8*: Row 1 – Proportion of sites where forecast performance (as measured by CRPS-S) is approximately equal

11 between root-zone soil-moisture and runoff. Row 2 – Proportion of sites for which root-zone soil-moisture or runoff

12 maximises forecast performance. Row 3 – The median CRPS-S forecast skill for sites where either root-zone soil-

13 moisture or runoff maximises forecast performance.

14

## 4. Discussion

### 4.1 Fixed input predictor configurations

The BJP predictor variable configurations defined in this study allows for an assessment of:

- Forecast performance based on observed catchment conditions only (benchmark model), where hydrological persistence alone determines the level of forecast skill at the catchment scale.

- Forecast performance of AWRA-L without an additional constraint from observed local catchment conditions relative to a defined benchmark model.

- Relative forecast performance from specific AWRA-L hydrological variables.

- Forecast performance from AWRA-L with input forcing from ACCESS-S2 while accounting for local catchment conditions (inclusion of antecedent observed streamflow – Qobs).

Low-flow season

The inclusion of observed antecedent conditions (Qobs) has greatest benefit in the low-flow period of the water-year. AWRA-L itself generates inter-annual variation that is either limited in magnitude and/or uncorrelated with the observed streamflows. This is most noticeable with runoff.

The fall in forecast performance occurs within the NAUS and SWC regions. Much of northern Australia (excluding the NEC region) experiences a tropical dry season whereas SWC has a strong Mediterranean climate (hot, dry summers). Both regions experience a particularly dry low-flow period. Other regions also experience a reduction in forecast skill. Somewhat surprisingly, Vic-SA exhibits a fall in skill, whereas adjacent regions of MDB-Sth and Tasmania exhibit a smaller reduction. The Vic-SA region may have more pronounced east-west hydroclimatic and geomorphological variation (the Victorian Alps existing in the eastern section of Vic-SA region). The north-east coast of Queensland (NEC region) exhibits comparatively small reduction in skill despite lying in northern Australia. This coastal region receives reliable rainfall even in the 'dry-season' due to persistent easterly trade-winds off the Coral Sea.

The low-flow season exhibits the biggest divergence in the proportion of sites that have positive forecast skill relative to the benchmark model. Across Australia, positive forecast skill is maintained for approximately 60% of sites maintain positive skill where Qobs is excluded. With Qobs included, the proportion jumps to ~75% (similar to the benchmark model). However, this difference is most apparent for NAUS, SWC and Vic-SA. Of note is the proportion of sites with very high forecast skill (e.g. ~50% sites in NAUS and 20% sites in SWC have CRPS-S skill > 50%), a result of very strong hydrological persistence (high autocorrelation). It is perhaps unreasonable to expect a process-based model, either calibrated nation-wide or at the catchment scale, to achieve such forecast skill without incorporating Qobs in some form to account for such strong autocorrelation (either through error-correction or data assimilation).

Approximately 20% of sites across Australia have negative forecast skill (CRPS-S < -5%), despite inclusion of Qobs, perhaps related to model bias that is not fully corrected for by the BJP model (where one would assume a forecast skill would revert to ~0 as a lower bound, equivalent to the historical climatology). A negative forecast skill may highlight violations of assumptions underlying the BJP model itself and limits of the data transformations used to

1    stabilize the variance of the BJP variables.

2    High-flow season

3    In contrast to the low-flow season, AWRA-L performs well compared to the benchmark model when Qobs is excluded

4    as an input predictor variable. The only noticeable exception is SWC. All other regions indicate either equivalent or

5    improved skill. This is true for median differences in forecast skill (Figure 5, Figure 6) as well as when comparing

6    the full distribution of forecast skill (Appendix A*1*, Appendix A*2*, Appendix A*3*, Appendix A*5*). The proportion of

7    sites with positive skill does not change compared to the benchmark model.

8    Inclusion of Qobs further improves forecast skill across Australia and is particularly noticeable in the MDBSth, Vic-

9    SA, NEC, NAUS (north Australian wet-season). Regions where there is little difference to the benchmark model

10   include Tasmania, MDBNth and coastal NSW (North and South). The significance of this result lies in the attribution

11   of skill improvement to ACCESS-S2 input forcing, as opposed to catchment initial state, the constraint for which is

12   provided by Qobs at the catchment scale. The same constraint representing local catchment conditions is used for the

13   benchmark model and influences forecast skill through hydrological persistence alone.

14   **4.2  Variable input variable configurations by site-month pairs**

15   The results presented in this paper demonstrate that either (root-zone) soil-moisture or runoff consistently maximises

16   forecast skill at individual site-month pairs as well as across different regions at different times of the year. A small

17   positive shift in the distribution of forecast skill (CRPS-S) occurs when the AWRA-L variable for which CRPS-S is

18   maximised is identified for each site-month pair (as represented by MAXSCR and MAXSCR-Q). Improved forecast

19   skill from either AWRA-L variable is a direct result of increased correlation with observed streamflow. A physical

20   explanation is improved representation of a hydrological process (represented either by soil-moisture or runoff) that

21   gives rise to an observed hydrological response (observed streamflow).

22   Overall, soil-moisture tends to maximise forecast skill during the low-flow season of the year (e.g. Vic-SA), although

23   an interesting exception is NAUS. Conversely, in the high-flow season there is an equal proportion of sites for which

24   either runoff or soil-moisture maximises forecast skill across Australia. This suggests that preference for either runoff

25   or soil-moisture is catchment specific or exists at a spatial scale smaller than the spatial regions defined in this study.

26   The difference between soil-moisture and runoff is instead a function of forecast performance itself (represented by

27   CRPS-S). The greater the forecast skill, the more likely runoff will maximise forecast skill relative to soil-moisture.

28   As an example, for the MDBSth region in the high-flow season (Appendix A4):

29       -    For soil-moisture, a greater proportion of sites have positive forecast skill (CRPS-S > 0)

30       -    For runoff, a greater proportion of sites have CRPS-S > 20%

31   The difference between root-zone soil-moisture and runoff is most clearly demonstrated when the respective CRPS-

32   S distributions are plotted separately for sites where either runoff or soil-moisture maximises forecast skill (Appendix

33   A4). This is apparent for most regions across Australia and for different times of the year. An exception is SWC.

34   **4.3  Strengths and Limitations of the AWRA-L model**

1  The analysis presented in this study highlights strengths and limitations of the AWRA-L hydrological model. If
2  AWRA-L generates inter-annual variation in either runoff or soil-moisture that is sufficient in magnitude and
3  correlated with observed streamflow, a BJP model calibrated against a fixed historical streamflow time-series is
4  sufficient to generate acceptable forecast skill. The BJP model correction is a function of the magnitude of the input
5  variables only. Lack of sufficient inter-annual variation necessitates an additional constraint correlated with observed
6  streamflow (i.e. observed antecedent observations).

7  For many site-month pairs, a statistical model calibrated on a fixed historical streamflow time-series is sufficient and
8  suggests AWRA-L is adequately representing the evolving catchment conditions with time (as represented by both
9  soil-moisture and runoff). No additional constraint at the catchment scale such as antecedent streamflow observations
10  is necessary (defined as Qobs(t-1) in Table 6).

11  The ability of AWRA-L to forecast streamflow appears largely driven by whether precipitation is a dominant input
12  forcing that elicits a hydrological response in both soil-moisture and runoff. AWRA-L performs well in the high-flow
13  period of the year. In the low-flow season for certain regions across Australia, AWRA-L cannot generate sufficient
14  inter-annual variation in either runoff or soil-moisture that can be used to forecast streamflow. This occurs
15  predominantly in regions where perhaps precipitation ceases to be a dominant input forcing. Other hydrological
16  processes, either missing in AWRA-L or simply parameterised rather than explicitly modelled, drive inter-annual
17  variability in streamflow during the low-flow season. Such variability exhibits strong hydrological persistence at the
18  monthly time-scale, enabling antecedent observed streamflow (Qobs) to provide a highly effective constraint to
19  forecast low-flows.

20  The extent to which precipitation elicits a hydrological response is different for runoff and soil-moisture due to
21  differences in hydrological persistence. The degree to which this persistence matches a response apparent in historical
22  streamflow observations determines the relative forecast performance of either soil-moisture or runoff. As forecast
23  performance falls, runoff is impacted more relative to soil-moisture, indicating that soil-moisture maintains a level of
24  hydrological persistence necessary to match the observed hydrological response (streamflow) at the monthly temporal
25  scale. This is consistent with Sharmila & Hendon, 2020 who identified soil-moisture as a good predictive indicator at
26  a similar time-scale due to hydrological persistence. Soil-moisture thus acts as a 'fallback option' when forecast skill
27  from AWRA-L reduces. In catchments where AWRA-L performs well, runoff will maximise forecast skill. When
28  performance reduces, soil-moisture maximises forecast skill. This is perhaps consistent with the finding of Pickett-
29  Heaps et al. (2021) & Vogel et al. (2021) that the seasonal forecast skill of root-zone soil-moisture is greater compared
30  to runoff across Australia. This improved forecast skill helps maintain a level of forecast skill for streamflow
31  comparable to the benchmark model, in an environment where forecast performance is otherwise limited. Forecast
32  skill in soil-moisture does not result in improved forecast skill of streamflow skill at longer lead-times (The MDBSth
33  region may be an exception). Instead, soil-moisture helps maintain an acceptable level of forecast skill for a greater
34  number of forecast locations across Australia during the first forecast month.

35  AWRA-L does not include a routing component in the calculation of accumulated runoff. This omission is
36  inconsequential for most uninterred catchments in this study (Figure 4) as routing is significant at the daily timescale
37  for the timing of flow peaks (Lerat et al., 2020), while aggregation to the monthly timescale reduces the impact of

1  omitting river routing. Moreover, inclusion of river routing is unlikely to remove the need for an error post-processing

2  at the monthly timescale.

### 4.4 The importance of antecedent observations to forecast streamflow

4  The inclusion of antecedent streamflow observations to maintain forecast skill is required in specific circumstances.

5  However, in general inclusion is not a strict dependency to achieve acceptable forecast skill (equivalent skill to the

6  benchmark model). Inclusion of antecedent streamflow observations as an input predictor variable might then consider

7  the following:

8  - Data latency – The reliable and timely delivery of hydrological observations when needed.

9  - Forecast publication – The option for issuing a forecast earlier than the forecast start-date (1st of the month).

10    Dependency on Qobs at the monthly temporal scale implies publication following the 1st of the month.

11  - Additional forecast locations – Forecasts at locations for which only historical hydrological data are available

12    only (e.g. temporary/permanent closure of a gauging station).

13  - Qobs classified as an optional predictor variable, included if it is available (i.e. data latency is acceptable).

14  - Inclusion of Qobs is site-dependent, depending on a positive impact on forecast skill and low data latency.

15  - Qobs redefined as aggregated streamflow observations up to a certain day of the month.

16  - The importance of individual forecast locations.

17  - Significance of hydrological flows in the low-flow period. Low flows may represent a small proportion to

18    annual flows or exhibit a small coefficient of variation.

19  Across Australia, the same proportion of sites achieve positive forecast skill for three out of four seasons of the water-

20  year (rising-limb, high-flow and falling-limb season) irrespective of whether Qobs is included as an input predictor

21  variable. There is a small difference in the proportion of sites with positive skill in the low-flow season. This difference

22  is primarily attributable to the SWC and NAUS regions in the low-flow season.

### 4.5 Impact of AWRA-L parameter optimization on validation results

24  Together with other observation datasets, gauged streamflow from a selection of catchments across Australia were

25  included in the optimisation of the AWRA-L model (version 6.1), resulting in a single parameter dataset. The

26  independence of any one of these gauging stations on the optimization of AWRA-L, if used in a verification study

27  such as this, is not completely guaranteed. However, given the large number of gauging stations selected for the

28  optimisation of AWRA-L, the impact on the assumption of independence of any single gauging station from the

29  AWRA-L optimisation is likely to be small.

30  To assess the possible impact on the verification results presented in this study, a relatively small subset of catchments

31  (denoted as subset A) included in this study were identified as having been included in the calibration dataset of the

32  AWRA-L model optimization (version 6.1). The remaining catchments are denoted as subset B. The forecast skill

33  distributions of each subset were then compared. The median forecast skill of subset A was found to be significantly

34  *lower* than that of subset B, whereas the opposite would be expected if independence from AWRA-L model

35  optimisation were to have a significant impact on the verification results presented in this study.

**5. Conclusion**

The major findings of this study are listed as follows:

- With statistical post-processing, AWRA-L can generate acceptable forecast skill at the seasonal (3-month) timescale.
- In general, inclusion of (observed) antecedent catchment conditions at the catchment scale need not be a strict requirement to ensure acceptable forecast skill.
- Inclusion of observed antecedent catchment conditions has the greatest benefit in the low-flow period of the water-year in specific regions across Australia.
- Inclusion of observed antecedent catchment conditions such as soil moisture can have an added benefit at other periods during the water-year.
- The spatio-temporal variation in relative forecast skill between runoff and root-zone soil-moisture explains strengths and limitations of the AWRA-L hydrological model.
- In general, when precipitation provides a dominant input forcing that elicits a hydrological response within AWRA-L, forecast performance at the monthly time-scale will increase.
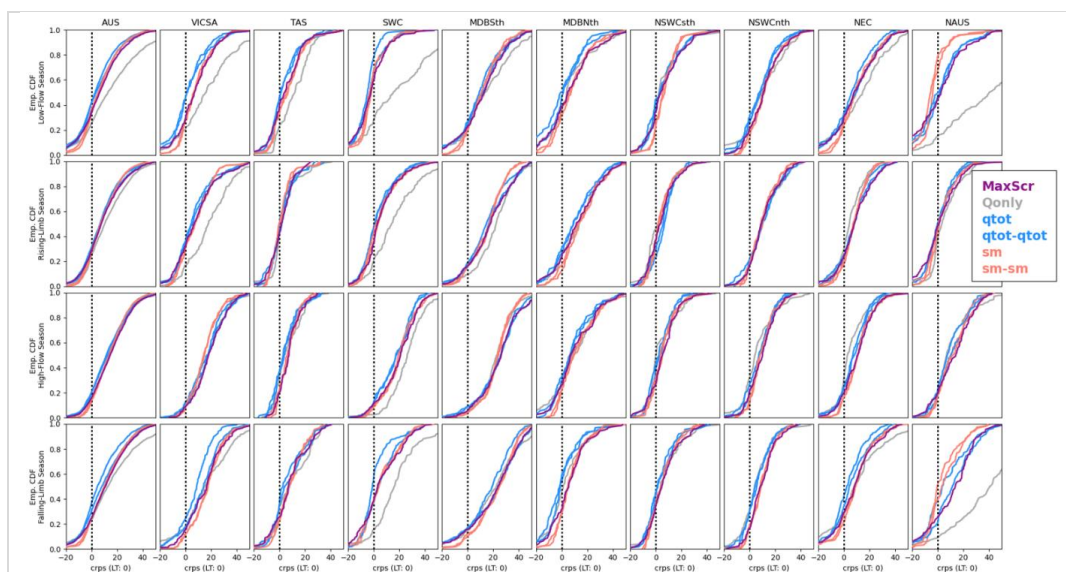
## 1    Appendix A        CRPS-S Empirical Cumulative Distribution Function Plots
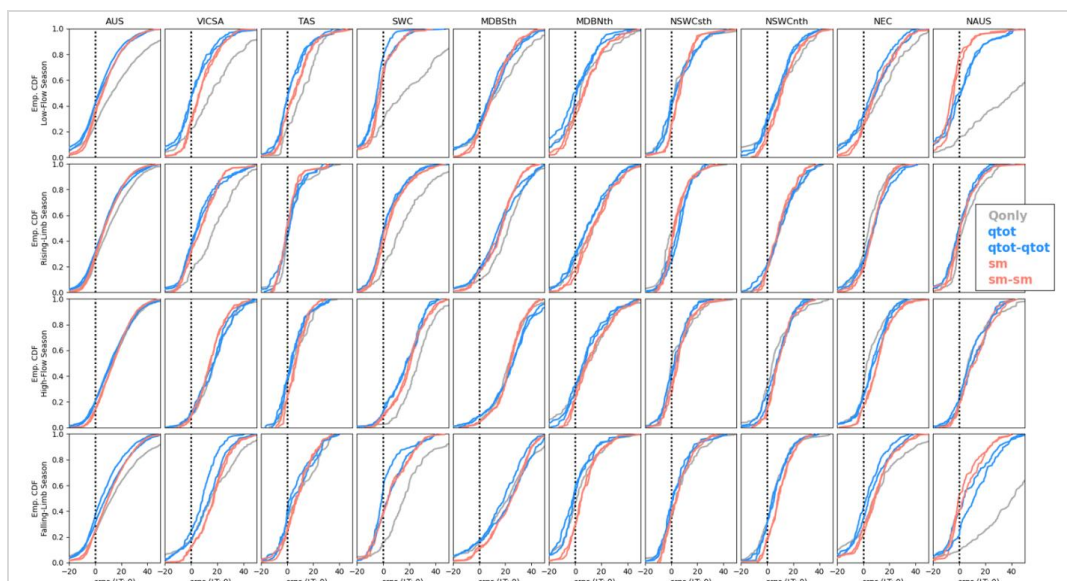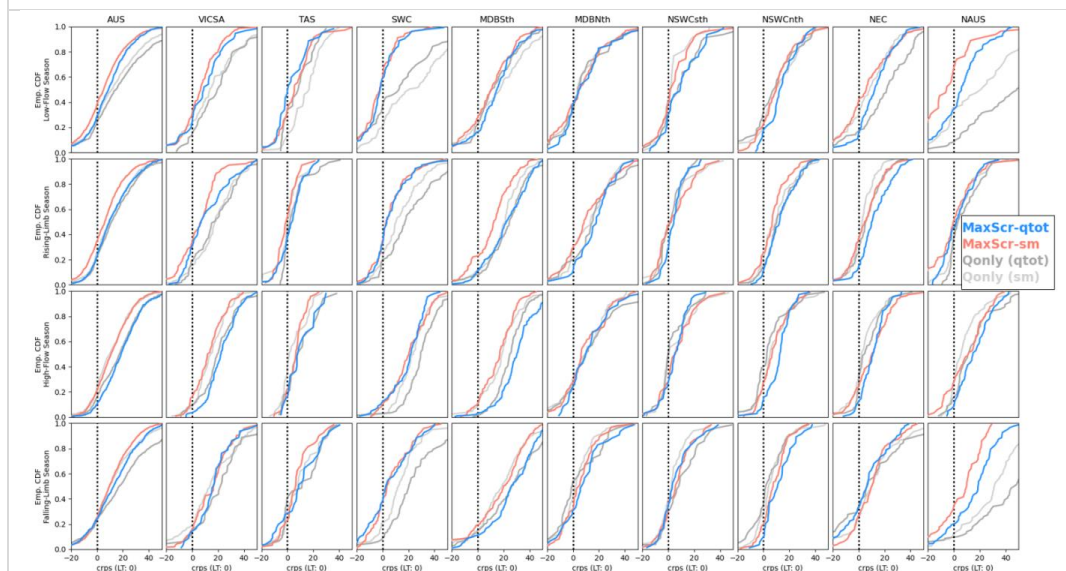


Appendix A1: CDF plots of CRPS-S forecast skill for AWRA-L input variables of runoff (qtot) and root-zone soil-moisture (sm). Benchmark model (Qonly) is also shown.
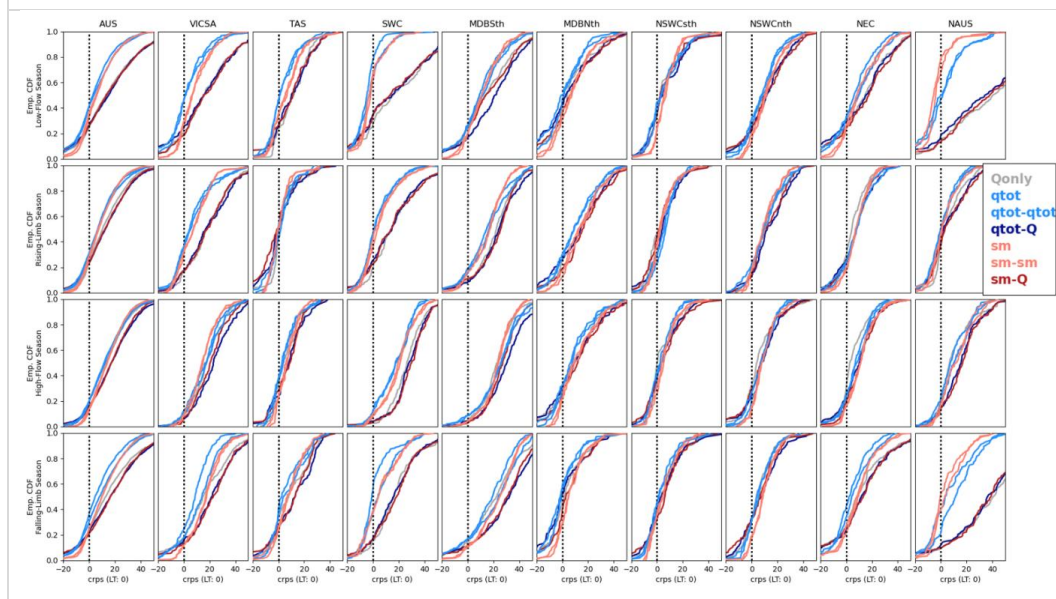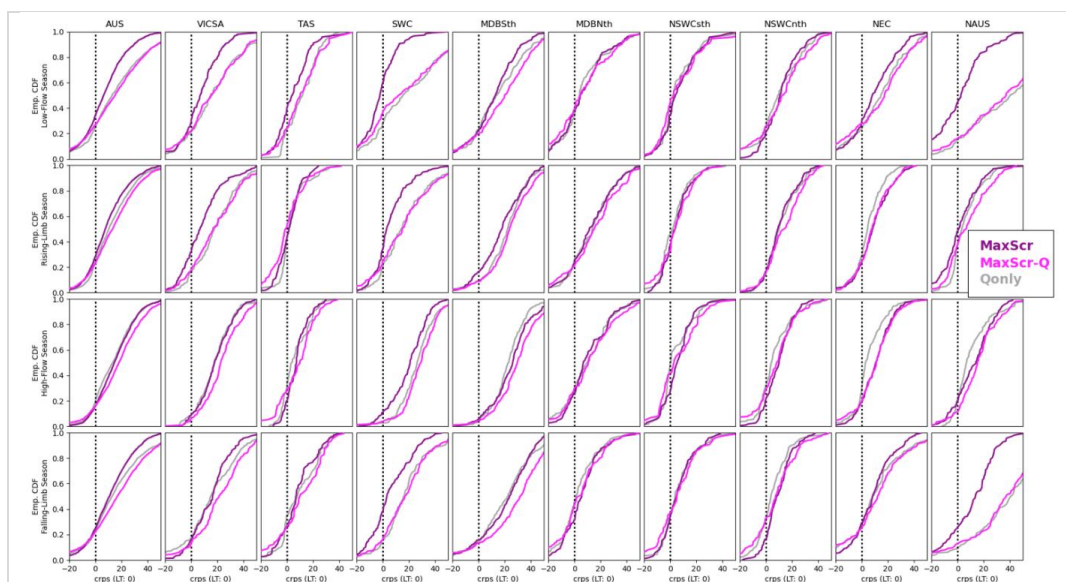


2    Appendix A2: CDF plots of CRPS-S forecast skill for AWRA-L input variables of runoff (qtot) and root-zone soil-

3    moisture (sm) either combined with Qobs or excluding Qobs. Benchmark model (Qonly) is also shown.

Appendix A3: CDF plots of CRPS-S forecast skill for AWRA-L input variables of runoff (qtot), root-zone soil-moisture (sm) and MaxScr (maximum possible forecast skill from either runoff (qtot) or root-zone soil-moisture (sm).



Appendix A4: CDF plots of CRPS-S forecast skill across the sample of sites where runoff (qtot) and root-zone soil-moisture (sm) maximise forecast skill. The CDF of the benchmark model (Qonly) for the same sample of sites is also included.

1

Appendix A5: CDF plots of CRPS-S forecast skill for both MaxScr and MaxScr-Q at lead-time 0. The benchmark model (Qonly) is also included.
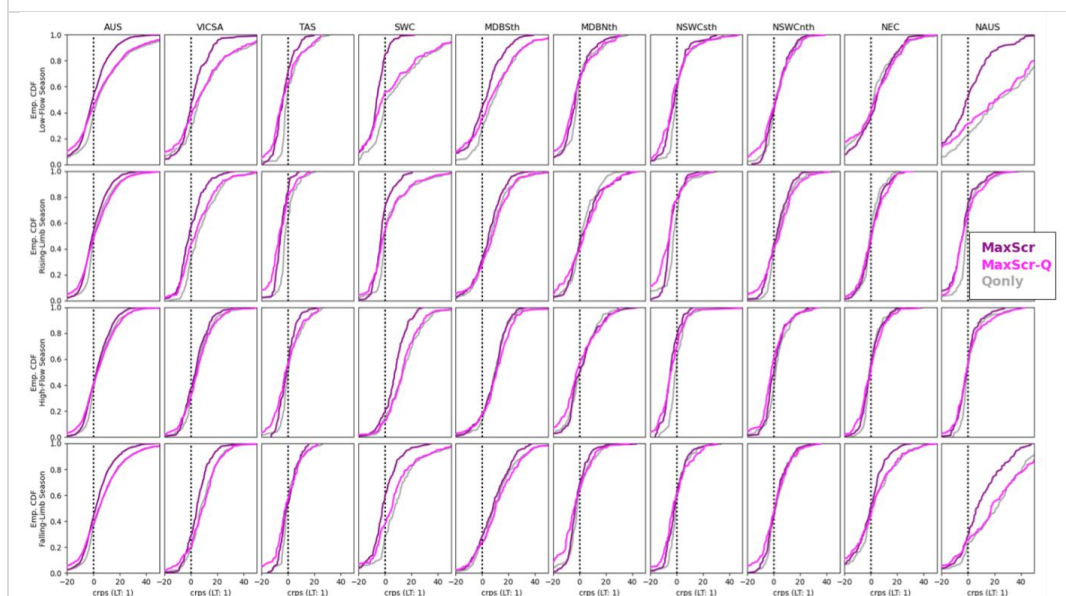


Appendix A6: CDF plots of CRPS-S forecast skill for both MaxScr and MaxScr-Q at lead-time 1. The benchmark model (Qonly) is also included.
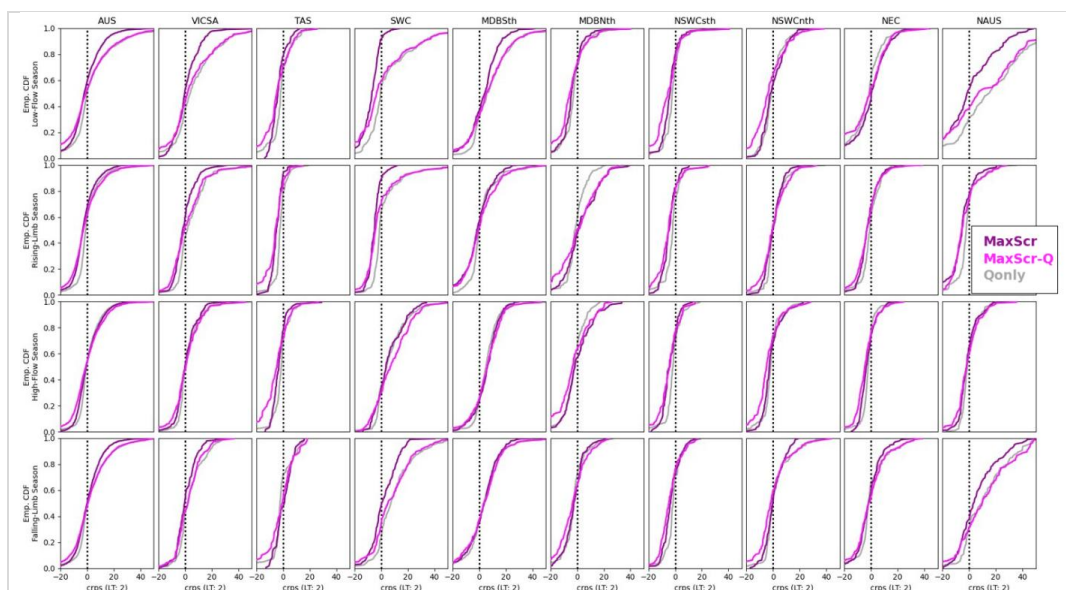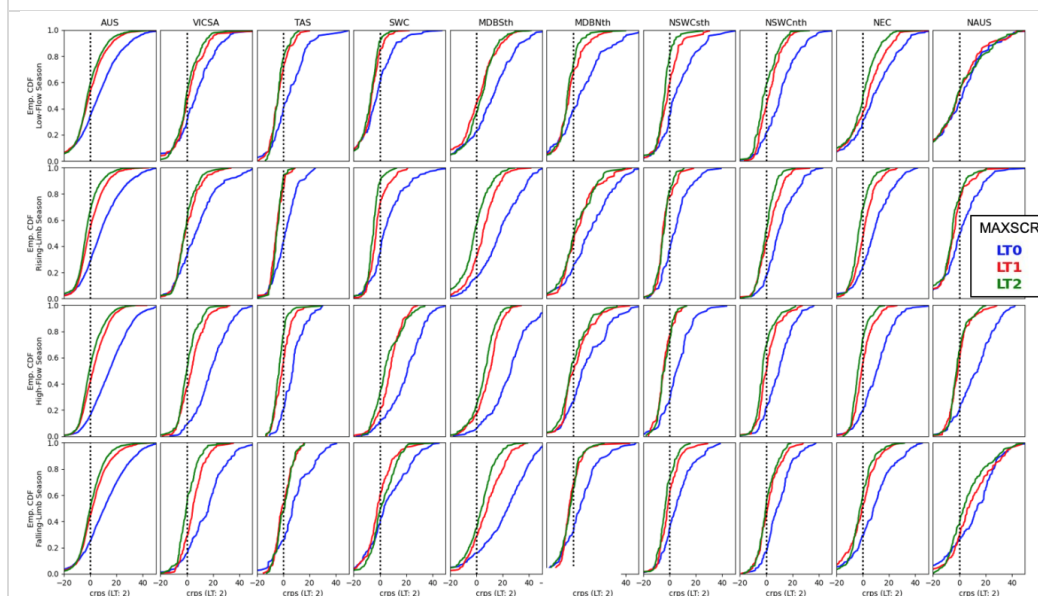
1

Appendix A7: CDF plots of CRPS-S forecast skill for both MaxScr and MaxScr-Q at lead-time 2. The benchmark model (Qonly) is also included.



Appendix A8: CDF plots of CRPS-S forecast skill for both MaxScr at lead-times 0, 1 & 2.

1

1 **Appendix B    Glossary**

2 The following definitions are relevant to this manuscript:

3 • **ACCESS-S**: Australian Community Climate and Earth-System Simulator – Seasonal ((Wedd et al., 2022)

4 • **(Areal) Actual Evapotranspiration (AET, etot)**: Sum of evaporation and transpiration (Vogel et al. 2021)

5 • **Australian Bureau of Meteorology (The Bureau)**: Australian federal government agency providing
6 meteorological, climatological and hydrological forecast and data products across Australia

7 • **Australian Gridded Climate Data (AGCD)**: < state definition >

8 • **Australian Water Outlook (AWO)**: A hydrological modelling and forecasting service provided by The
9 Bureau:

10 • **Bayesian Joint Probability model (BJP)**: A statistical model based on a multivariate Gaussian distribution:
11 awo.bom.gov.au

12 • **BJP month-site pair**: An individual BJP model defined for a particular month at a particular forecast location
13 (encompassing all monthly lead-times).

14 • **Forecast lead-time** refers to a monthly forecast period relative to the seasonal forecast issue date, as defined
15 in Table B1.

| Forecast Month | 1st | 2nd | 3rd | Month prior |
|---|---|---|---|---|
| Lead-time | 0 | 1 | 2 | -1 |
| B1. Lead-time –1: Month prior to seasonal forecast issue date (representing a period for which antecedent conditions/recent observations are available) | | | | |

16 • **Forecast location**: Either a river gauging station or total inflow site

17 • **High-flow period (water-year)**: The 6-month period in which high-flows occur at any given site across
18 Australia during the locally defined water year. The 6-month high-flow season encompasses the months 6-
19 11 of the water-year.

20 • **High-flow season (water-year)**: The 3-month high-flow season encompasses the months 7-9 of the locally
21 defined water-year.

22 • **Hindcast**: A retrospective forecast (i.e. a forecast run in the past, usually for comparisons with concurrent
23 observations

24 • **Hindcast dataset**: A set of retrospective forecasts within a specified period (e.g. 1981-2017)

25 • **HPC**: High performance computation

26 • **Hydrologic Reference Stations (HRS)** dataset: A dataset of high-quality hydrological observations across
27 Australia (Amirthanathan et al., 2023): www.bom.gov.au/water/hrs/

28 • **Low-flow period (water-year)**: The 6-month period in which low-flows occur at any given site across
29 Australia during the locally defined water year. The 6-month low-flow season encompasses the months 1-5
30 and month 12 of the water-year.

31 • **Low-flow season (water-year)**: The 3-month low-flow season encompasses the months 1-3 of the locally
32 defined water-year.

1     •  **Lead-up to the high-flow season (water-year)**: A 3-month period of the water year leading up to the defined
2         3-month high-flow season, generally encompassing the months 4-6 of the water-year.

3     •  **MCMC sampling**: Markov Chain Monte Carlo sampling

4     •  **New South Wales (NSW):** A state jurisdiction in eastern Australia

5     •  **Potential Evapotranspiration ($e_0$):** Upper-limit on AET and controlled by the available energy at the earth
6         surface as well as water availability.

7     •  **Runoff (qtot):** The combination of surface runoff, baseflow and interflow (Vogel et al. 2021).

8     •  **Seasonal forecast:** individual monthly forecasts extending out to 3 months (seasonal timescale). Seasonal
9         forecasts (and respective monthly lead-times) are always relative to the 1st of the month of the forecast issue
10         date.

11     •  **Seasonal Streamflow Forecasting (SSF) service**: A streamflow forecasting service provided by The
12         Bureau: www.bom.gov.au/water/ssf

13     •  **Soil-moisture (*sm*)**: Refers to root-zone soil-moisture (see section **Error! Reference source not found.**,
14         Table *10*). Total soil water available within the 0-1.0m soil depth. Expressed as absolute soil-moisture (mm)
15         and as the fraction of available water holding capacity of the soil column (smpfull).

16     •  **Streamflow (Qobs)**: Gauged streamflow (or river discharge) observations at the monthly timescale (unless
17         stated otherwise)

18     •  **Water-year**: A 12-month year referenced to seasonality in hydroclimate observations (either rainfall or
19         streamflow), where the first season (3-month period) refers to the 'low-flow' season (i.e. the start of the
20         water-year coincides with the month in which the lowest streamflow is observed). The water-year is locally
21         defined, either at a given site (gauging station) or region. Given opposing seasonality in precipitation across
22         Australia, the calendar month that coincides with the start of the water-year varies across the country.

23 **Appendix C     Bayesian Joint Probability Statistical Model**

24 The theory presented as follows is provided for context and is reproduced from T. Zhao et al., 2016. The BJP model
25 defines a multi-variate Gaussian distribution (*P(X)*) of a set of random variables (*X*), the solution of which provides
26 the 'joint' probability (P) of all possible combinations across *X*.

| | | |
|---|---|---|
| $P(X) \sim N(\hat{\mu}, \hat{\Sigma})$ | *1* | •  $\hat{\mu}$ is a vector of (expected) values (mean) across the random variables in $X$ |
| $z \sim N(\hat{\mu}, \hat{\Sigma})$ | *2* | •  $\Sigma$ is the covariance matrix |
| $\hat{\Sigma} = \hat{\sigma}.R.\hat{\sigma}^T$ | *3* | •  $z^T = [z_1, z_2, z_3, \dots z_d]$ for d scalar random variables in $X$<br>•  $\hat{\sigma}^2$ is a vector of (estimated) variance ($[\sigma_1^2, \sigma_2^2, \sigma_3^2, \dots \sigma_d^2]$)<br>•  $R$ is the correlation coefficient matrix |

27 T. Zhao et al., 2016

28 In the context of this study, the random scalar variables in *X* are defined as follows for a single forecast location:

29     •  AWRA-L forecasts of either root-zone soil-moisture or runoff for the defined catchment upstream of a single

1         forecast location (gauging station)

2       •     Observed streamflow at the forecast location

3   These variables are defined at specific lead-times. The marginal distribution of each random variable $x$ in $X$ is

4   $\sim N(\hat{\mu}, \hat{\sigma})$. The parameterized covariance matrix allows for an estimation of the variance and covariance between all

5   random variables in $X$.

6   Random variables $\mathbf{X}$ can be split into two groups:

| | |
|---|---|
| $X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ <br><br> $z^T = \begin{bmatrix} z_1^T \\ z_2^T \end{bmatrix}$ | ▪   $\mathbf{X_1}$ is defined as a set of predictor variables <br><br> ▪   $\mathbf{X_2}$ is defined as a set of predictand variables |

7   T. Zhao et al., 2016

8   By fixing a subset of random variables in $X$ (i.e. $z_1$) to known values (i.e. AWRA-L seasonal forecasts at specified

9   lead-times), one can then solve for the *conditional probability* of the remaining unknown random variables ($z_2$) as

10   follows by first partitioning $\hat{\boldsymbol{\mu}}$ and $\hat{\Sigma}$:

| | | |
|---|---|---|
| $\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$ | *4* | Partitioning of $\boldsymbol{\mu}$ |
| $\Sigma = \begin{bmatrix} \Sigma_{1,1} & \Sigma_{1,2} \\ \Sigma_{2,1} & \Sigma_{2,2} \end{bmatrix}$ | *5* | Partitioning of the covariance matrix: $\Sigma$ |
| $p(\boldsymbol{z_2}\vert\boldsymbol{z_1}) \approx N(\boldsymbol{\mu'}_2, \Sigma'_{2,2})$ | *6* | Conditional probability of $z_2$ given $z_1$ |
| $\boldsymbol{\mu'}_2 = \mu_2 + \Sigma_{2,1}\left[\Sigma_{1,1}\right]^{-1}\left[z_1 - \mu_1\right]$ | *7* | Updated expected value $\boldsymbol{\mu'}_2$ based on the conditional probability $p(\boldsymbol{z_2}\vert\boldsymbol{z_1})$ |
| $\Sigma'_{2,2} = \Sigma_{2,2} - \Sigma_{2,1}\left[\Sigma_{1,1}\right]^{-1}\Sigma_{1,2}$ | *8* | Updated estimate of $\Sigma'_{2,2}$ based on the conditional probability $p(\boldsymbol{z_2}\vert\boldsymbol{z_1})$ |

11   Zhao et al. 2016

12   Positive correlation in $\boldsymbol{R}$ results in the conditional probability distribution of $X_2$ as having reduced variance relative to

13   the marginal distribution of $X_2$. The conditional probability also results in a bias correction ($\boldsymbol{\mu'}_2$) that is a function of

14   $\boldsymbol{\mu_1}$. The conditional probability $N(\boldsymbol{\mu'}_2, \Sigma'_{2,2})$ can then be sampled to generate realisations of the unknown random

15   variables in $X_2$. A single-valued (or deterministic) solution of the BJP model is obtained by calculating the mean or

16   median from the sampled conditional probability. Similarly, confidence intervals (5,95 percentiles) can also be

17   obtained from the sampled conditional probability.

18    In a practical sense, the BJP model provides:

1       • A bias correction as a function of the constraints provided by fixed random variables in $X_1$

2       • An estimation of the uncertainty in the updated estimates of $X_2$ (reduced relative to that of the marginal

3       distribution of $X_2$).

4    The BJP model is beneficial for operational hydrological modelling/forecasting in two ways:

5       • There is little impact of missing 'predictor' data for variables assigned to $X_1$ (those with a fixed value). If

6       certain predictor variables remain unknown (due to missing data), these scalar random variables can simply

7       be assigned to $X_2$, for which a conditional probability can be calculated. The model does not fail due to the

8       presence of missing input data, possibly resulting in a missing forecast. This simplifies the modelling system

9       underpinning BJP.

10      • With no correlation between variables in $X_1$ and $X_2$, the conditional probability of variables in $X_2$ reverts to

11      their respective marginal distributions. In the context of hydrological forecasting, variables in $X_2$ assigned to

12      greater lead-times will have reduced correlation with predictor variables of $X_1$ at the same lead-times.

13      • The BJP modelling package used in this study is robust and well-tested, having supported the SSF service

14      since its inception in 2010.

15    *Data transformations*

16    The joint probability *P(X)* assumes that all data follow a Gaussian distribution. Hydrological variables, either from

17    AWRA-L or observed in the environment, are heteroskedastic – Variance with respect to magnitude (e.g. streamflow

18    magnitude) is unstable and results in highly skewed distributions. Data transformations can be applied to stabilize the

19    variance so that it approximates a Gaussian distribution (e.g. Woldemeskel et al., 2018). Following T. Zhao et al.

20    (2016) and Wang et al., 2012, the *log-sinh* data transformation is applied to all hydrological data used in the calibration

21    and application of a BJP model and is defined as follows:

| | | |
|---|---|---|
| $z = \dfrac{1}{\lambda} \log\left(sinh([\varepsilon + \lambda \boldsymbol{y}])\right)$ | | Log-sinh transformation |
| $y = \dfrac{arcsinh(e^{\lambda z}) - \varepsilon}{\lambda}$ | | Inverse transformation |
| $y$ | | Original hydrological data time-series |
| $z$ | | Transformed data |

22                                                  Zhao et al. (2015)

23    The log-sinh transformation was developed specifically for hydrological and precipitation data (Wang et al., 2012).

24    The transformation parameters ($\lambda$, $\varepsilon$) are optimized independently at each gauging station using the available historical

25    hydrological data time-series.

26    *Censored data*

27    In addition to exhibiting heteroskedasticity, streamflow data in arid regions are bounded by a lower threshold. This

28    threshold (T) is typically 0, indicating zero or otherwise immeasurable streamflow. Wang & Robertson, 2011 provide

29    a means to account for observed data with a lower bound T by treating such data as *censored data* when calibrating a

1    BJP model. The BJP implementation used in this study includes the methodology of BJP calibration described in

2    Wang et al. (2011) that treats hydrological data bounded by a threshold (T=0) as censored data.

3    ***BJP model calibration***

4    The application of a BJP statistical model requires a historical time-series of all random variables defined in $\mathbf{P(X)}$ over

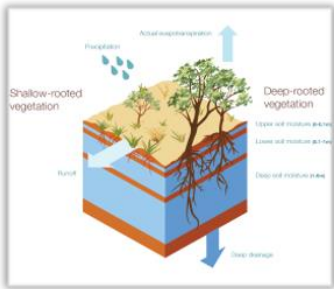5    a common time-period and consistent temporal frequency.

6

7

1     **Appendix D     Grid-Based Modelling Systems**

2        **The Australian Water Resource and Assessment Landscape Model (AWRA-L)**

3 The AWO is underpinned by the Australian Resource and Assessment Landscape model (AWRA-L). AWRA-L is a

4 landscape model (or alternatively a soil-moisture model) that simulates the spatial/temporal dynamics of the Australian

5 surface water balance (table D1), as represented by.

6      •     The temporal evolution of hydrological stores: soil-moisture three separate soil-layers in the soil-column.

7      •     The temporal evolution of hydrological fluxes that result from the evolution of hydrological stores (which in

8        turn are influence by input forcing: section **Error! Reference source not found.** and **Error! Reference**

9        **source not found.**).

10 AWRA-L is defined on a $5km^2$ (0.05deg) spatial grid across the Australian continent, where each $5km^2$ grid-cell

11 represents a single soil-column. AWRA-L is run at a daily time-step, requiring all input climate forcing to be provided

12 at the same daily time-step. AWRA-L consists of two natural hydrological response units (HRU) representative of

13 shallow-rooted and deep-rooted vegetation. Processes not included in AWRA-L include (1) lateral flows between

14 adjacent grid-cells and (2) river routing. This study uses AWRA-L version 6.1. A thorough description of the AWRA-

15 L model structure and calibration is given in Frost et al., 2018 and a comprehensive evaluation is provided by Frost

16 & Write (2018).

| Hydrological stores | Depth | Variable | Unit | |
|---|---|---|---|---|
|     Upper Layer | 0 – 0.1m | s0 | mm | |
|     Lower Layer | 0.1 – 1.0m | ss | mm | |
|     Root-zone layer[1] | 0 – 1.0m | sm | mm | |
|     Deep Layer | 1.0 – 6.0m | sd | mm | |
| **Hydrological fluxes** | | | | |
|     Runoff | | qtot | mm | |
|     Actual ET[3] | | etot | mm | |
|     Potential ET[3] | | e0 | mm | |
|     Deep Drainage | | dd | mm | |
| **Climate input forcing** | | | | Schematic source: AWO & Frost et al., 2018 |
|     Precipitation | | p | mm | |
|     Daily min/max temperature | | $t_{max}/t_{min}$ | C | |
|     Daily short-wave solar radiation | | w | $W.m^{-2}$ | |
|     Daily mean surface wind | | | $m.s^{-1}$ | |
| | | | | |
| 1: Root-zone soil layer: Derived from the sum of the upper layer & lower layer | | | | |
| 2: Soil-moisture in any layer can be expressed as % of available capacity for the defined soil layer | | | | |
| 3: ET: Evapotranspiration | | | | |

1    Table 91: The Australian Water Resources and Assessment (land) model – AWRA-L

2    **AWO historical simulations**

3    AWO provides a historical simulation daily from 1911 to the present, updated daily to provide the status of the

4    Australian water balance. Daily climate forcing is provided by the Australian Gridded Climate dataset (AGCD: Evans

5    et al., 2020; Jones et al., 2009). The service also includes data assimilation of surface soil-moisture (defined as the

6    soil-moisture in the top 2cm of the soil-column) retrieved from two satellite instruments (Tian et al., 2021).

7    **AWO seasonal forecasts**

8    The AWO provides seasonal forecasts (see section **Error! Reference source not found.**) of the Australian water

9    balance as represented by four key AWRA-L variables (Table *10***Error! Reference source not found.**D2, Pickett-

10    heaps & Vogel, 2022; Vogel et al., 2021).

| AWO forecast variable | Variable ID | Unit | Temporal resolution | Forecast duration |
|---|---|---|---|---|
| Root-zone soil-moisture | sm | %full[1] | | |
| Runoff | qtot | mm | 1-month | 3-months |
| Actual ET | etot | mm | | |
| Potential ET[2] | $e_0$ | mm | | |
| 1: SM provided as a % of available capacity in the defined soil-column | | | | |
| 2: Not publicly available | | | | |

11    Table 10: Key hydrological variables of the Australian water balance, for which AWRA-L generates seasonal

12    forecasts.

13    The sole difference between AWRA-L historical simulations and seasonal forecasts is the input forcing. Climate

14    forcing for seasonal forecasts is provided by the ACCESS-S (vS2) seasonal climate model. Unless otherwise stated,

15    soil-moisture refers to root-zone soil-moisture.

16    **ACCESS-S climate model (vS2)**

17    The Australian Community Climate Earth System Simulator – Seasonal (ACCESS-S, Wedd et al., 2022) is a multi-

18    week to seasonal ensemble-based climate forecasting system consisting of:

19    •    A coupled atmosphere-ocean-land Earth-system model

20    •    A data assimilation system

21    •    An ensemble generation system

22    ACCESS-S2 (denoting version S2) runs at a spatial resolution of 60km$^2$ at an hourly time-step out to 6 months.

23    Forecast ensembles represent a set of equally likely forecast outcomes (or realisations), enabling the probability of an

24    event to be quantified. ACCESS-S2 is configured to run daily, generating 11 ensembles per day. A full seasonal

25    forecast consists of 99 lagged ensembles: 9 sets of 11 ensembles extending back 9 successful daily forecast runs (i.e.

26    failed runs are not 'restarted'). A large ensemble (99 members) ensures the probability of rare events is captured by

27    the probabilistic forecast, despite a degradation of forecast skill with ensembles that have greater ensemble lag.

1     Computational constraints limit the number of ensembles that can be run each day.

2     **ACCESS-S2 calibrated forecasts**

3     Raw ACCESS-S2 forecast output are calibrated (i.e. downscaled/bias-corrected) to a 5km2 spatial resolution,

4     consistent with the AGCD and AWRA-L grid specification, by applying quantile-quantile (Q-Q) mapping (Morwenna

5     et al. 2023, BOM 2019). The reference dataset for this Q-Q model is the AGCD dataset (1981-2018). Calibrated

6     ACCESS-S2 forecasts are statistically consistent with AGCD climate grids (1981-2018), with which the Q-Q model

7     is calibrated. Consequently, AWRA-L simulations with either AGCD or ACCESS-S input forcing (individual

8     ensembles) are equivalent over the long-term (1981-2018).

9     **Appendix E     Forecast Skill Metrics**

10    *Cumulative-Rank-Probability-Score (CRPS)*

11    CRPS is defined as follows (Hersbach, 2000):

$$CRPS_i = C(x) = \int_{-\infty}^{\infty} [P(x) - P_a(x)]^2 . dx$$

$$\overline{CRPS} = \sum_{i=1}^{N} w_i . C(x_i)$$

$$P(x) = \int_{-\infty}^{x} p(y). dy$$

$$P_a(x) = H(x - x_a)$$

$$H(x) = \begin{cases} 0 \ for \ x \ < 0 \\ 1 \ for \ x \ \geq 0 \end{cases}$$

- $CRPS_i$ is the CRPS value for forecast *i* out of N forecasts in a hindcast dataset
- *P(x)* is the cumulative distribution function of the forecast PDF *p(x)* of a forecast variable *x*
- $x_a$ is the observed value of an event forecast by *p(x)*
- *$P_a(x)$* is the cumulative distribution of the observation, following a Heaviside function (*H(x)*)
    - H(x) is effectively a step function where the cumulative probability shifts from 0 to 1 instantaneously at *x=$x_a$*.
- $\overline{CRPS}$ is the mean/weighted-mean CRPS value across all available N forecasts.

12

13    Whereas other forecasts metrics can only be applied to a deterministic forecast, CRPS can be applied to ensemble

14    forecasts. CRPS considers forecast ensemble spread (i.e. precision) as well as forecast bias (accuracy). When applied

15    to a deterministic forecast, CRPS is equivalent to the mean absolute error (Hersbach, 2000).

16    Like other skill metrics, CRPS is a statistical quantity. Bootstrapping sampling was used to generate a median estimate

17    of CRPS for each site-month pair and the uncertainty in the CRPS estimate (the analysis of which is not included in

18    this manuscript). Examples of CRPS uncertainty in hydrological forecasting at individual forecasting locations are

19    available from the [SSF operational service](#).

20    *Forecast reliability*

1 Forecast reliability (Wilks, 2019) relates forecast probabilities to observed frequencies for a set of forecast values,
2 where ideally the relationship is 1:1. More generally, reliability relates to forecast ensemble consistency, where an
3 observed outcome is statistically equivalent to and indistinguishable from any forecast ensemble. A forecast model is
4 reliable if the distribution of the rank of each observation within a forecast ensemble is approximately uniform across
5 all forecasts. The Kolmogorov-Smirnov (KS) test is used to statistically verify if observed ranks relative to forecast
6 ensembles are uniformly distributed, with a p-value statistic and threshold of <0.05 used to identify if a model is
7 unreliable.

8 Forecast reliability does not itself indicate a high level of forecast skill. Instead, a forecast model must maintain
9 forecast reliability while increasing forecast sharpness (precision). This implies improved forecast accuracy (a
10 reduction in model absolute bias) relative to a reference model.

11 ***Other forecast metrics***

12 Other forecast metrics considered in this study include:

- Correlation
- Relative bias
- Forecast sharpness (Woldemeskel et al., 2018)
- Tercile hit rates.

13 **Author Contributions**

14 CPH – Conceptulization, methodology, investigation, software, validation & formal analysis, data curation, writing
15 full draft, vizulisation

16 PS – Conceptulization, software, project administration, resources

17 WS – Supervision, Resources, Writing – review and editing

18 MP – Software, resources

19 CW – Software, data curation

20 AC – Project administration, funding acquisition

21 RL – Project administration

22 EC - Supervision, Resources, Writing – review and editing

23 **Acknowledgements**

1 **References**

2     Amirthanathan, G. E., Bari, M. A., Woldemeskel, F. M., Tuteja, N. K., & Feikema, P. M. (2023). Regional

3         significance of historical trends and step changes in Australian streamflow. *Hydrology and Earth System*

4         *Sciences*, *27*(1). https://doi.org/10.5194/hess-27-229-2023

5     Ashok, K., Guan, Z., & Yamagata, T. (2003). Influence of the Indian Ocean Dipole on the Australian winter rainfall.

6         *Geophysical Research Letters*, *30*(15). https://doi.org/10.1029/2003GL017926

7     Bennett, J. C., Wang, Q. J., Pokhrel, P., & Robertson, D. E. (2014). The challenge of forecasting high streamflows

8         1-3 months in advance with lagged climate indices in southeast Australia. *Natural Hazards and Earth System*

9         *Sciences*, *14*(2). https://doi.org/10.5194/nhess-14-219-2014

10     Boer, M. M., Resco de Dios, V., & Bradstock, R. A. (2020). Unprecedented burn area of Australian mega forest

11         fires. In *Nature Climate Change* (Vol. 10, Issue 3). https://doi.org/10.1038/s41558-020-0716-1

12     Brown, L. D., Cai, T. T., & Das Gupta, A. (2001). Interval estimation for a binomial proportion. *Statistical Science*,

13         *16*(2). https://doi.org/10.1214/ss/1009213286

14     Chiew, F. H. S., Piechota, T. C., Dracup, J. A., & McMahon, T. A. (1998). El Nino/Southern Oscillation and

15         Australian rainfall, streamflow and drought: Links and potential for forecasting. *Journal of Hydrology*, *204*(1–

16         4). https://doi.org/10.1016/S0022-1694(97)00121-2

17     Devanand, A., Falster, G. M., Gillett, Z. E., Hobeichi, S., Holgate, C. M., Jin, C., Mu, M., Parker, T., Rifai, S. W.,

18         Rome, K. S., Stojanovic, M., Vogel, E., Abram, N. J., Abramowitz, G., Coats, S., Evans, J. P., Gallant, A. J.

19         E., Pitman, A. J., Power, S. B., … Ukkola, A. M. (2024). Australia's Tinderbox Drought: An extreme natural

20         event likely worsened by human-caused climate change. *Science Advances*, *10*(10).

21         https://doi.org/10.1126/sciadv.adj3460

22     Evans, A., Jones, D., Smalley, R., & Lellyett, S. (2020). An enhanced gridded rainfall dataset scheme for Australia.

23         In *Bureau Research Report No. 41* (Issue June).

24     Feikema, P. M., Wang, Q. J., Zhou, S., Shin, D., Robertson, D. E., Schepen, A., Lerat, J., Bennett, J. C., Tuteja, N.

25         K., & Jayasuriya, D. (2018). Service and Research on Seasonal Streamflow Forecasting in Australia. *World*

26         *Scientific Series on Asia-Pacific Weather and Climate*, *10*. https://doi.org/10.1142/9789813235663_0010

27     Frost, Andrew J, Shokri, A. (2021). The Australian Landscape Water Balance model (AWRA-L v7). Technical

28         Description of the Australian Water Resources Assessment Landscape model version 7. *Bureau of*

29         *Meteorology Technical Report*. https://awo.bom.gov.au/assets/notes/publications/AWRA-

30         Lv7_Model_Description_Report.pdf

31     Frost, A. J. ., Ramchurn, A. ., & Smith, A. . (2018). The Australian Landscape Water Balance model (AWRA-L v6).

1      Technical Description of the Australian Water Resources Assessment Landscape model version 6. *Bureau of*
2      *Meteorology Technical Report*.
3      https://awo.bom.gov.au/assets/notes/publications/AWRALv6_Model_Description_Report.pdf

4      Ghajarnia, N., Arasteh, P. D., Araghinejad, S., & Liaghat, M. A. (2016). A hybrid Bayesian-SVD based method to
5      detect false alarms in PERSIANN precipitation estimation product using related physical parameters. *Journal*
6      *of Hydrology*, *538*. https://doi.org/10.1016/j.jhydrol.2016.04.037

7      Griffiths, M., Smith, P., Yan, H., Spillman, C., & Young, G. (2023). ACCESS-S2 : Updates and improvements to
8      postprocessing pipeline. In *Bureau Research Report No. 082* (Vol. 72, Issue May).

9      Hao, Z., Singh, V. P., & Xia, Y. (2018). Seasonal Drought Prediction: Advances, Challenges, and Future Prospects.
10     *Reviews of Geophysics*, *56*(1). https://doi.org/10.1002/2016RG000549

11     Hendon, H. H., Thompson, D. W. J., & Wheeler, M. C. (2007). Australian rainfall and surface temperature
12     variations associated with the Southern Hemisphere annular mode. *Journal of Climate*, *20*(11).
13     https://doi.org/10.1175/JCLI4134.1

14     Hersbach, H. (2000). Decomposition of the continuous ranked probability score for ensemble prediction systems.
15     *Weather and Forecasting*, *15*(5). https://doi.org/10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2

16     Holgate, C. M., Van Dijk, A. I. J. M., Evans, J. P., & Pitman, A. J. (2020). Local and Remote Drivers of Southeast
17     Australian Drought. *Geophysical Research Letters*, *47*(18). https://doi.org/10.1029/2020GL090238

18     Huang, A. T., Gillett, Z. E., & Taschetto, A. S. (2024). Australian Rainfall Increases During Multi-Year La Niña.
19     *Geophysical Research Letters*, *51*(9). https://doi.org/10.1029/2023GL106939

20     Jones, D. A., Wang, W., & Fawcett, R. (2009). High-quality spatial climate data-sets for Australia. *Australian*
21     *Meteorological and Oceanographic Journal*, *58*(4). https://doi.org/10.22499/2.5804.003

22     Lerat, J., Thyer, M., McInerney, D., Kavetski, D., Woldemeskel, F., Pickett-Heaps, C., Shin, D., & Feikema, P.
23     (2020). A robust approach for calibrating a daily rainfall-runoff model to monthly streamflow data. *Journal of*
24     *Hydrology*, *591*. https://doi.org/10.1016/j.jhydrol.2020.125129

25     Li, M., Wang, Q. J., Bennett, J. C., & Robertson, D. E. (2016). Error reduction and representation in stages (ERRIS)
26     in hydrological modelling for ensemble streamflow forecasting. *Hydrology and Earth System Sciences*, *20*(9).
27     https://doi.org/10.5194/hess-20-3561-2016

28     McInerney, D., Thyer, M., Kavetski, D., Laugesen, R., Tuteja, N., & Kuczera, G. (2020). Multi-temporal
29     Hydrological Residual Error Modeling for Seamless Subseasonal Streamflow Forecasting. *Water Resources*
30     *Research*, *56*(11). https://doi.org/10.1029/2019WR026979

1   McInerney, D., Thyer, M., Kavetski, D., Lerat, J., & Kuczera, G. (2017). Improving probabilistic prediction of daily
2       streamflow by identifying Pareto optimal approaches for modeling heteroscedastic residual errors. *Water
3       Resources Research*, *53*(3). https://doi.org/10.1002/2016WR019168

4   Pickett-heaps, C. A., & Vogel, E. (2022). Seasonal Hydrological Ensemble Forecasts for Australia using AWRA-L
5       – Hindcast Verification Report. In *Bureau Research Report No. 065* (Issue July).

6   Pokhrel, P., Robertson, D. E., & Wang, Q. J. (2013). A Bayesian joint probability post-processor for reducing errors
7       and quantifying uncertainty in monthly streamflow predictions. *Hydrology and Earth System Sciences*, *17*(2).
8       https://doi.org/10.5194/hess-17-795-2013

9   Risbey, J. S., Pook, M. J., McIntosh, P. C., Wheeler, M. C., & Hendon, H. H. (2009). On the remote drivers of
10      rainfall variability in Australia. *Monthly Weather Review*, *137*(10). https://doi.org/10.1175/2009MWR2861.1

11  Robertson, D. E., & Wang, Q. J. (2012). A Bayesian approach to predictor selection for seasonal streamflow
12      forecasting. *Journal of Hydrometeorology*, *13*(1). https://doi.org/10.1175/JHM-D-10-05009.1

13  Schepen, A., Zhao, T., Wang, Q. J., & Robertson, D. E. (2018). A Bayesian modelling method for post-processing
14      daily sub-seasonal to seasonal rainfall forecasts from global climate models and evaluation for 12 Australian
15      catchments. *Hydrology and Earth System Sciences*, *22*(2). https://doi.org/10.5194/hess-22-1615-2018

16  Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and Statistical Modeling with Python. *Proceedings of
17      the 9th Python in Science Conference*. https://doi.org/10.25080/majora-92bf1922-011

18  Sharmila, S., & Hendon, H. H. (2020). Mechanisms of multiyear variations of Northern Australia wet-season
19      rainfall. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-020-61482-5

20  Tian, S., Renzullo, L. J., Pipunic, R. C., Lerat, J., Sharples, W., & Donnelly, C. (2021). Satellite soil moisture data
21      assimilation for improved operational continental water balance prediction. *Hydrology and Earth System
22      Sciences*, *25*(8). https://doi.org/10.5194/hess-25-4567-2021

23  Ummenhofer, C. C., Gupta, A. Sen, Briggs, P. R., England, M. H., McIntosh, P. C., Meyers, G. A., Pook, M. J.,
24      Raupach, M. R., & Risbey, J. S. (2011). Indian and Pacific Ocean influences on southeast Australian drought
25      and soil moisture. *Journal of Climate*, *24*(5). https://doi.org/10.1175/2010JCLI3475.1

26  Van Dijk, A. I. J. M., Beck, H. E., Crosbie, R. S., De Jeu, R. A. M., Liu, Y. Y., Podger, G. M., Timbal, B., & Viney,
27      N. R. (2013). The Millennium Drought in southeast Australia (2001-2009): Natural and human causes and
28      implications for water resources, ecosystems, economy, and society. *Water Resources Research*, *49*(2).
29      https://doi.org/10.1002/wrcr.20123

30  Van Dijk, A. I. J. M., Peña-Arancibia, J. L., Wood, E. F., Sheffield, J., & Beck, H. E. (2013). Global analysis of
31      seasonal streamflow predictability using an ensemble prediction system and observations from 6192 small

catchments worldwide. *Water Resources Research*, *49*(5). https://doi.org/10.1002/wrcr.20251

Vogel, E., Lerat, J., Pipunic, R., Frost, A. J., Donnelly, C., Griffiths, M., Hudson, D., & Loh, S. (2021). Seasonal ensemble forecasts for soil moisture, evapotranspiration and runoff across Australia. *Journal of Hydrology*, *601*. https://doi.org/10.1016/j.jhydrol.2021.126620

Wang, Q. J., & Robertson, D. E. (2011). Multisite probabilistic forecasting of seasonal flows for streams with zero value occurrences. *Water Resources Research*, *47*(2). https://doi.org/10.1029/2010WR009333

Wang, Q. J., Robertson, D. E., & Chiew, F. H. S. (2009). A Bayesian joint probability modeling approach for seasonal forecasting of streamflows at multiple sites. *Water Resources Research*, *45*(5). https://doi.org/10.1029/2008WR007355

Wang, Q. J., Shrestha, D. L., Robertson, D. E., & Pokhrel, P. (2012). A log-sinh transformation for data normalization and variance stabilization. *Water Resources Research*, *48*(5). https://doi.org/10.1029/2011WR010973

Wang, Q. J., Zhao, T., Yang, Q., & Robertson, D. (2019). A seasonally coherent calibration (SCC) model for postprocessing numerical weather predictions. In *Monthly Weather Review* (Vol. 147, Issue 10). https://doi.org/10.1175/MWR-D-19-0108.1

Wedd, R., Alves, O., De Burgh-Day, C., Down, C., Griffiths, M., Hendon, H. H., Hudson, D., Li, S., Lim, E. P., Marshall, A. G., Shi, L., Smith, P., Smith, G., Spillman, C. M., Wang, G., Wheeler, M. C., Yan, H., Yin, Y., Young, G., … Zhou, X. (2022). ACCESS-S2: the upgraded Bureau of Meteorology multi-week to seasonal prediction system. *Journal of Southern Hemisphere Earth Systems Science*, *72*(3). https://doi.org/10.1071/ES22026

Wheeler, M. C., Hendon, H. H., Cleland, S., Meinke, H., & Donald, A. (2009). Impacts of the Madden-Julian oscillation on australian rainfall and circulation. *Journal of Climate*, *22*(6). https://doi.org/10.1175/2008JCLI2595.1

Wilks, D. S. (2019). Statistical Methods in the Atmospheric Sciences, Fourth Edition. In *Statistical Methods in the Atmospheric Sciences, Fourth Edition*. https://doi.org/10.1016/C2017-0-03921-6

Woldemeskel, F., McInerney, D., Lerat, J., Thyer, M., Kavetski, D., Shin, D., Tuteja, N., & Kuczera, G. (2018). Evaluating post-processing approaches for monthly and seasonal streamflow forecasts. *Hydrology and Earth System Sciences*, *22*(12). https://doi.org/10.5194/hess-22-6257-2018

Zhao, P., Wang, Q. J., Wu, W., & Yang, Q. (2022a). Extending a joint probability modelling approach for post-processing ensemble precipitation forecasts from numerical weather prediction models. *Journal of Hydrology*, *605*. https://doi.org/10.1016/j.jhydrol.2021.127285

1    Zhao, P., Wang, Q. J., Wu, W., & Yang, Q. (2022b). Spatial mode-based calibration (SMoC) of forecast

2        precipitation fields from numerical weather prediction models. *Journal of Hydrology*, *613*.

3        https://doi.org/10.1016/j.jhydrol.2022.128432

4    Zhao, T., Schepen, A., & Wang, Q. J. (2016). Ensemble forecasting of sub-seasonal to seasonal streamflow by a

5        Bayesian joint probability modelling approach. *Journal of Hydrology*, *541*.

6        https://doi.org/10.1016/j.jhydrol.2016.07.040

7    Zhao, T., Wang, Q. J., Bennett, J. C., Robertson, D. E., Shao, Q., & Zhao, J. (2015). Quantifying predictive

8        uncertainty of streamflow forecasts based on a Bayesian joint probability model. *Journal of Hydrology*, *528*.

9        https://doi.org/10.1016/j.jhydrol.2015.06.043

10

11