

Performance evaluation of the national 7-day water forecast service

H.A.P. Hapuarachchi ^a, A. Kabir ^a, X.S. Zhang ^a, D. Kent ^a, M.A. Bari ^b, N.K. Tuteja ^c, M. M. Hasan ^c,
D. Enever ^a, D. Shin ^a, K. Plastow ^a, Z. Ahmad ^a

^a Bureau of Meteorology, Water Forecasting Services, Melbourne

^b Bureau of Meteorology, Water Forecasting Services, Perth

^c Bureau of Meteorology, Water Forecasting Services, Canberra

Email: sophie.zhang@bom.gov.au

Abstract: The Bureau of Meteorology launched a deterministic 7-day streamflow forecast service to the public in September, 2015. These streamflow forecasts indicate likely river flow conditions in the coming week. The service has been developed mainly to assist river managers in making informed decisions about water resources management to ensure the best use is made, both from agricultural and environmental perspectives. The service currently provides daily updates of daily volume forecasts to the public for 132 locations across Australia (Figure 1). This paper investigates the performance of the operational 7-day streamflow forecasts at targeted 53 locations across Australia.

Evaluating operational forecast performance is essential to identify forecast discrepancies in order to improve the service. Performance metrics for rainfall and streamflow were investigated. Operational streamflow forecasts for more than 90% of the selected 53 locations

exceed a set criterion of positive mean absolute error skill score up to 3-days lead-time. Unreliable or poor quality input data was the main cause of suboptimal performance for the remaining 10% of locations. Forecast performance for perennial streams is found to outperform those in ephemeral streams. The inability of hydrological model structure to cope with complex and highly non-linear hydrological processes in dry areas is seem to be the cause for suboptimal streamflow forecasts. Forecast performance at locations with large catchment areas show better skills than other locations having small catchment areas.

Keywords: Water, forecasting, streamflow, 7-day, modeling, hydrology, Bureau of Meteorology

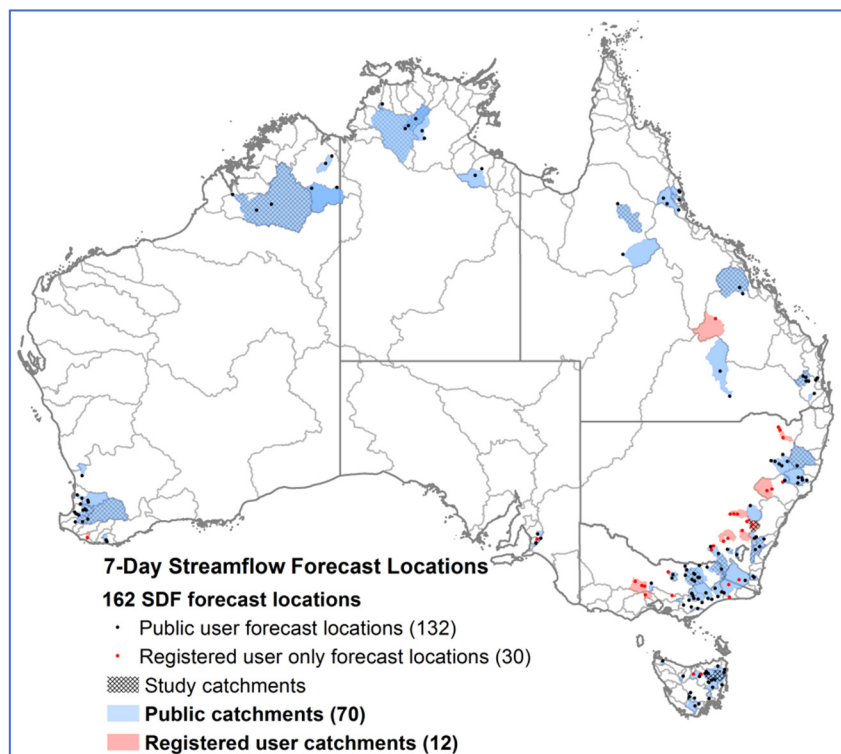


Figure 1. Study locations and catchments

1. INTRODUCTION

The Bureau of Meteorology launched a deterministic [7-day streamflow forecast](#) (SDF) service for Australia in September 2015 as part of its water information role and responsibilities under the [Water Act 2007](#). This is the first national operational continuous streamflow forecast system developed in Australia. These streamflow forecasts indicate likely river flow volumes in the coming week. The service was developed mainly to assist river managers in making informed decisions about water resources to ensure the best use is made of natural river inflows, both from agricultural and environmental perspectives. It was developed in partnership with key water agencies and the science underpinning this service is supported by CSIRO and Bureau's Science for Services section.

At present, the service provides daily streamflow forecasts to the public for 132 locations in 70 catchments across all jurisdictions in Australia (Figure 1). An additional 36 forecast locations are planned to be released to the public in November 2017. A customised hourly streamflow forecast product in addition to the daily forecast volume product is also available to [registered users](#), who mainly represent river managers, reservoir operators, and irrigators. The registered user service is available for 209 forecast locations in 100 catchments. Further information about the development of this service can be found in Hapuarachchi *et al.* (2016).

Quantifying performance of the operational forecasts is crucial to diagnose errors and in the planning of development work to improve forecast accuracy and extend the forecast lead time. This paper presents forecast evaluation results for 53 selected locations within the operational 7-day streamflow forecast service.

2. OPERATIONAL SYSTEM

The Bureau of Meteorology's next generation Hydrological Forecasting System (HyFS) is used to generate 7-day streamflow forecasts operationally. HyFS is the central national platform for modelling that underpins flood forecasting and warning activity in Australia. HyFS is a Delft-FEWS (Flood Early Warning System) based forecasting environment (see <http://oss.deltares.nl/web/delft-fews/about>). This system allows for management of input observations and numerical weather prediction (NWP) model rainfall forecasts, input data processing, forecasting and maintenance workflows, model internal state management and forecast visualization. The core hydrological modeling package in the forecast system is the Short-term Water Information Forecasting Tools (SWIFT, Ward *et al.*, 2012; Pagano *et al.*, 2011). SWIFT is a streamflow modelling package designed for both operational streamflow forecasting and scientific research. It is a collection of hydrologic modelling components and utilities that support model calibration, validation, and forecast verification. For the SDF service, the GR4H model (hourly), which is an adaptation of the GR4J (Perrin *et al.*, 2003) conceptual rainfall runoff model, is used based on the evidence of good performance in Australia from previous research. A catchment is delineated to small subareas (100-800km²) and GR4H is applied to each subarea. Then the runoff is routed to the catchment outlet using Muskingum channel routing (Cunge 1969). Simulated streamflow is post-processed using the dual pass error correction scheme (Pagano *et al.* 2011). Real-time gauged rainfall and streamflow data to run SWIFT operationally are prepared by the HyFS system. An inbuilt workflow in HyFS is used for data quality checking. Hourly streamflow forecasts of up to 7 days' lead-time are generated once a day using Australian Community Climate and Earth-System Simulator ([ACCESS-G](#)) rainfall forecasts. These forecasts are fed into an SDF product generator to produce plots, tables and data files, and the output is published in a web portal (www.bom.gov.au/water/7daystreamflow). The whole process is fully automated and forecasts are updated daily between 10:00AM and 11:00AM AEST.

3. STUDY CATCHMENTS

Evaluation of rainfall and streamflow forecasts was carried out for 53 forecast locations in 23 catchments across all regions where the SDF service is operational. These catchments were selected to cover a wide range of sizes, climate conditions, and hydrological characteristics (Figure 1 and Table 1).

4. DATA AND METHODOLOGY

4.1. Data

Real-time input data and output data of the operational service are archived daily since the start of the service. For this study, archived observed and forecast data generated by the operational system were used for evaluating the forecast performance for the period October 2015 to June 2017 (21 months). Although there is an inbuilt quality checking procedure in the HyFS, we found very poor quality input data at two selected forecast locations due to the observation sensor malfunctioning for a short period. These data were

removed from the evaluation procedure. In practice when a sensor issue is identified, the particular forecast location is disabled from the service until the issues is resolved.

Table 1. Selected forecast locations for evaluating the SDF performance

Region	No.	Station Name	Catchment Area (km ²)	Region	No.	Station Name	Catchment Area (km ²)
NSW	1	Abercrombie River at Hadley No.2	1,630	SA	28	Onkaparinga D/S Hahndorf Creek	224
	2	Abercrombie River at Abercrombie No.2	2,632		29	Florentine River above Derwent River	445
	3	Tantawangalo Creek at Candelo Damsite	201	TAS	30	Florentine River at Eleven Road	169
	4	Bemboka River at Morans Crossing	316		31	St. Pauls River above Avoca	520
	5	Cotter River at Gingera	130		32	South Esk River at Llewellyn	2,287
	6	Goobarragandra River at Lacmalac	668	VIC	33	Fifteen Mile Creek at Greta South	224
	7	Hunter River at Belltrees	1,189		34	Rose River at Matong North	179
	8	Hunter River at Moonam Dam Site	742		35	Buffalo River at Abbeyard	420
	9	Tia River at Tia	265		36	Ovens River at Bright	490
	10	Shoalhaven River at Fossickers Flat	4,660		37	Buckland River at Harris Lane	457
	11	Shoalhaven River at Hillview	2,661		38	Ovens River at Myrtleford	1,158
	12	Shoalhaven River at Warri	1,376		39	Ovens River at Rocky Point	2,964
	13	Murray River at Biggara	1,257		40	Ovens River at Wangaratta	5,138
NT	14	Adelaide River at Railway Bridge	638		41	Ovens River at Peechelba East	6,187
	15	Katherine River near Birdie Creek	3,624		42	Seven Creeks at Kialla West	1,519
	16	Katherine River at Nitmiluk Centre	7,293		43	Stoney Creek at Tamleugh	370
	17	Katherine River at Katherine Bridge	8,371	44	Seven Creeks at Euroa	346	
QLD	18	Cooyar Creek at Taromeo Ck	956	WA	45	Blackwood River at Boyup Brook Flax Mill	16,556
	19	Brisbane River at Linville	1,999		46	Blackwood River at Bridgetown	17,767
	20	Emu Creek at Boat Mountain	928		47	Blackwood River at Nannup	19,201
	21	Brisbane River at Gregor Creek	3,875		48	Collie River at Buckingham Mill	792
	22	Gilbert River at Rockfields	11,537		49	Collie River East at Coolangatta Farm	1,328
	23	Stanley River at Peachester Alert	102		50	Collie River at South Branch	666
	24	Stanley River at Woodford Alert-B	246		51	Fitzroy River at Fitzroy Crossing Bridge	45,696
	25	Tully River at Euramo	1,388		52	Fitzroy River at Noonkanbah	65,619
	26	Connors River at Pink Lagoon	8,701		53	Fitzroy River at Willare	83,154
	27	Isaac River at Yatton	19,673				

4.2. Rainfall Evaluation

Semi-distributed hydrological models use subarea average rainfall: the mean areal rainfall estimated to fall within that subarea. However, in many cases, ACCESS-G (APS2) grid cells (25km) are larger than subareas. Therefore, in this study we evaluated the average rainfall in the upstream area of each forecast location. The average rainfall (for both observations and forecasts) was computed by integrating area-weighted subarea rainfall over the upstream area of each forecast location. Gauged rainfall at each subarea centroid was derived by inverse distance weighting of rainfall from nearby rain gauges. Rainfall forecasts were downscaled to subareas by taking the area-weighted average of gridded forecast rainfall for all grid cells intersecting the subarea. Forecasts were evaluated at a daily time step. This required aggregating 3-hourly ACCESS-G forecasts to daily totals. Area average forecast rainfall was compared with observed values of up to 7 days' lead-time. The forecasts of each lead-time (1 – 7 days) of the 21-month period were then evaluated using the metrics given in Section 4.4.

4.3. Streamflow Evaluation

Operational streamflow forecasts archived from October 2015 to June 2017 (21 months) were compared with observed values for forecast locations up to 7 days' lead-time. Daily mean streamflow was calculated using the hourly streamflow observations and forecasts. Forecast evaluation was carried out similar to the rainfall evaluation.

4.4. Forecast Evaluation Metrics

Evaluating operational forecast performance is useful to identify forecast discrepancies in order to improve the service. River operators are mainly interested in unbiased cumulative volume of water contributing to a reservoir or the main channel. Also they wanted to see the forecast skill with reference to a certain threshold. To accommodate these, the following metrics were used to evaluate the performance and forecast quality of models on a daily basis. The same metrics were used for the rainfall and streamflow evaluations.

4.4.1 Bias

It is important to assess model bias to ensure the model is not consistently underestimating or overestimating streamflow. Bias (*Bias*) can be positive (underestimation) or negative (overestimation), and was calculated for each lead time (*LT*) using:

$$Bias_{LT} = \frac{\sum_{i=1}^n (Q_{i,obs,LT} - Q_{i,sim,LT})}{\sum_{i=1}^n (Q_{i,obs,LT})} \quad (1)$$

where $Q_{i,obs}$ is observed streamflow, $Q_{i,sim}$ is simulated streamflow, LT is lead-time, and n is total number of observations.

4.4.2 Cumulative Volume Error (%) – (CVE)

Cumulative Volume Error for each lead time LT (CVE_{LT}) was computed as:

$$CVE_{LT} = \frac{\sum_{i=1}^n \sum_{LT=1}^{LT} (Q_{i,obs,LT} - Q_{i,sim,LT})}{\sum_{i=1}^n \sum_{LT=1}^{LT} (Q_{i,obs,LT})} \quad (2)$$

4.4.3 Mean Absolute Error (MAE)

Mean Absolute Error (MAE) is the average of the magnitude of the errors. The perfect score is zero, and was calculated by:

$$MAE_{LT} = \frac{\sum_{i=1}^n |Q_{i,obs,LT} - Q_{i,sim,LT}|}{n} \quad (3)$$

4.4.4 Mean Absolute Error Skill Score (MAESS)

Skill Score (SS) is a measure of forecast accuracy with respect to the reference forecast accuracy.

A skill score based on MAE is known as the Mean Absolute Error Skill Score (MAESS). It is calculated as:

$$MAESS_{LT} = 1 - \frac{\sum_{i=1}^n |Q_{i,obs,LT} - Q_{i,sim,LT}|}{\sum_{i=1}^n |Q_{i,obs,LT} - Q_{i,obs,c}|} \quad (4)$$

where $Q_{i,obs,c}$ is climatological streamflow calculated over the period from 1995 to 2014. For any given day of the year the climatology value will be the median of the period from 2 weeks before that day to 2 weeks after (i.e. 29 days) over the climatology period excluding the forecast year.

4.4.5 Nash-Sutcliffe model Efficiency coefficient (NSE)

The Nash-Sutcliffe efficiency (*NSE*) quantifies the relative magnitude of residual variance compared to the measured data variance, by:

$$NSE_{LT} = 1 - \frac{\sum_{i=1}^n (Q_{i,obs,LT} - Q_{i,sim,LT})^2}{\sum_{i=1}^n (Q_{i,obs,LT} - \bar{Q})^2} \quad (5)$$

where \bar{Q} is mean observed streamflow.

NSE values may range from $-\infty$ to 1. The higher values imply greater accuracy. It can be easily compared across different catchments. For 7-day streamflow forecast performance, all the models with $NSE \geq 0.6$ were considered to be satisfactory.

5. RESULTS AND DISCUSSION

Figure 2 shows overall forecast performance for the 53 locations evaluated in this study. It shows the number of forecast locations not exceeding the absolute bias (%), absolute cumulative volume error (%), and the mean absolute error (mm) scores at different lead times for rainfall and streamflow.

Generally, streamflow bias increases with lead-time. About 70% of the catchments have forecast rainfall MAE less than 2mm at 1-day lead-time, while the streamflow MAE is less than 0.25mm (Figure 2: a3 and b3). This is mainly due to streamflow post-processing which has an effect up to 3-4 days lead time. Streamflow MAE of the rest of the 30% of forecast locations increases rapidly. Most of these are ephemeral catchments with upstream contributing areas less than 4,000 km². The relative bias becomes large when the denominator is small (equation 1) particularly in ephemeral catchments. Forecast rainfall bias is less than 10% only for 20% of forecast locations for 1-day lead-time. This implies that forecast rainfall error largely contributes to the suboptimal streamflow forecasts, and that streamflow post-processing has a large impact on

forecast performance particularly at shorter lead times. Although it is generally observed that the quality of forecast rainfall decreases with lead time, there is no much variation in CVE and bias here (Figure 2: a1 and a2). This is mainly because CVE and bias represent relative values. Therefore MAE is a better metric to compare the rainfall quality with lead-time.

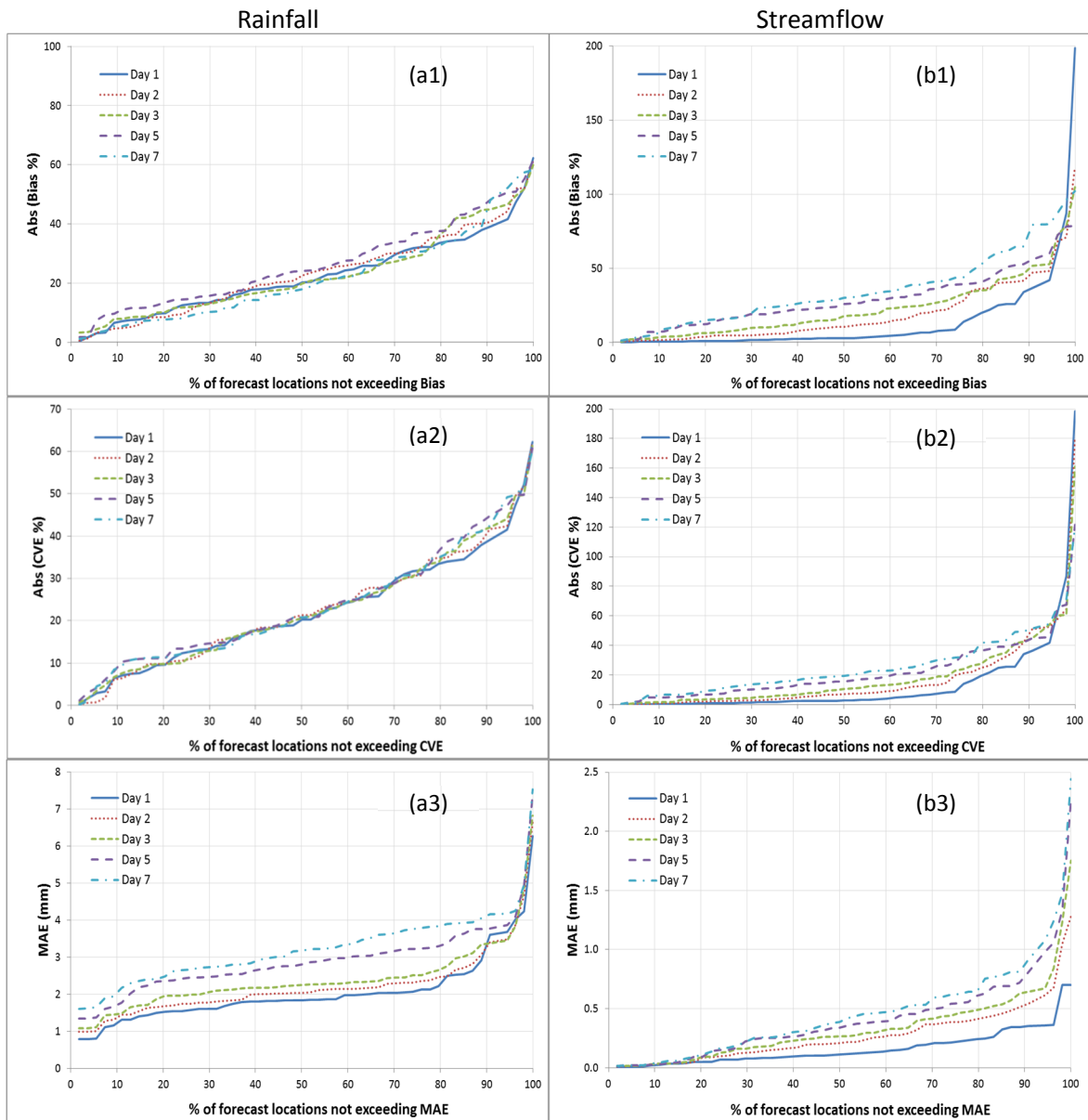


Figure 2. Forecast evaluation metrics for catchment average forecast rainfall (a1-a3) and streamflow forecasts (b1-b3) for lead-times 1-day to 7-days.

Figure 3 shows the streamflow forecast skill (NSE) at each forecast location up to 7 days' lead-time. The NSE score for most locations is satisfactory up to 7 days' lead-time. In most catchments of NSW and QLD, the 21 months data we used had significant low-flow or no-flow periods, and has only a few large events. It was found that the magnitude of these large events control the overall forecast skill. During the evaluation period, streamflow forecasts in the Brisbane River catchment (Table 1: No. 18-21) show poor skills at all forecast locations, although the rainfall forecast skill was reasonably good. The forecast streamflow volume for April-May 2017 was much higher value when compared to observed which greatly reduces the NSE. This particular rainfall event happened just after three consecutive rainfall events that brought the model soil moisture state to its full capacity. Nonexistence of such a rainfall pattern in the model calibration period (2007-2014) led the catchment model having relatively small soil moisture and ground water capacity than

what it should be in reality. Therefore the model forecasted substantial overland flow, when in reality more rainfall infiltrated into the soil.

On the other hand, streamflow forecasts at the Katherine River catchment (Table 1, No. 15-17) produces reasonably good forecast skill while the rainfall forecast shows considerably poor skill. Gradual changes in river flows in the Katherine River catchment driven by large contributions from groundwater, results in reasonably good streamflow forecasts during the evaluation period. The underpinning catchment characteristics, the flow persistence, and the respective rainfall pattern during the period led to good model performance.

Figure 4 shows the forecast skill (MAESS) at 3-days lead-time at the 53 forecast locations. A positive MAESS up to 3-days lead-time has been used as the criterion for releasing a forecast location to public. Most forecast locations show good forecast skill irrespective of climate region. However, some locations in ephemeral catchments (temperate climate) in South Australia, Western Australia, north western part of Victoria, and some parts of inland New South Wales have relatively lower skill. Overall, about 10% of locations do not satisfy the criterion, although all locations satisfied the criterion when evaluated offline using a long period (2011-2014) of hindcast data (results present in the [SDF website](#)). The observed input data used in offline evaluation was carefully quality checked. One reason for lower operational forecast performance at a very small number of forecast locations is low quality real-time input data (observed and forecast). As a result, hydrological model initial states calculated using observed data (first 4 days) were erroneous in some instances, and the error was propagated to the streamflow forecasts. In addition, the errors in a few large events have dominated the forecast skill score due to relatively small (21 months) evaluation data sample size.

The differences in forecast skill were due to nonlinear dynamic behavior in catchment hydrologic processes, particularly the interaction between groundwater and channel flow. Estimation of streamflow for a relatively dry or ephemeral catchment is not consistent through a conceptual (lumped) modelling approach due to the simple approximation of surface and groundwater hydrologic processes. It is worth noting that all SDF catchment models underwent a thorough cross-validation process and produced satisfactory results during the hindcast period (from 2011 to 2014). However, discrepancies in rainfall distribution and pattern, and antecedent soil moisture conditions could lead to poor streamflow forecasts. Depending on prevailing climate and hydrological conditions, a catchment could behave as perennial, intermittent, or dry during certain periods. More research is necessary to improve conceptual hydrological models to underpin these highly variable natural conditions.

6. CONCLUSIONS

We found operational streamflow forecasts for over 90% of the targeted 53 locations exceed a set criterion of positive Mean Absolute Error Skill Score (MAESS) up to 3-days lead-time. However, all locations passed the criterion when evaluated offline with hindcast data for a longer period dataset (results shown in the [SDF](#)

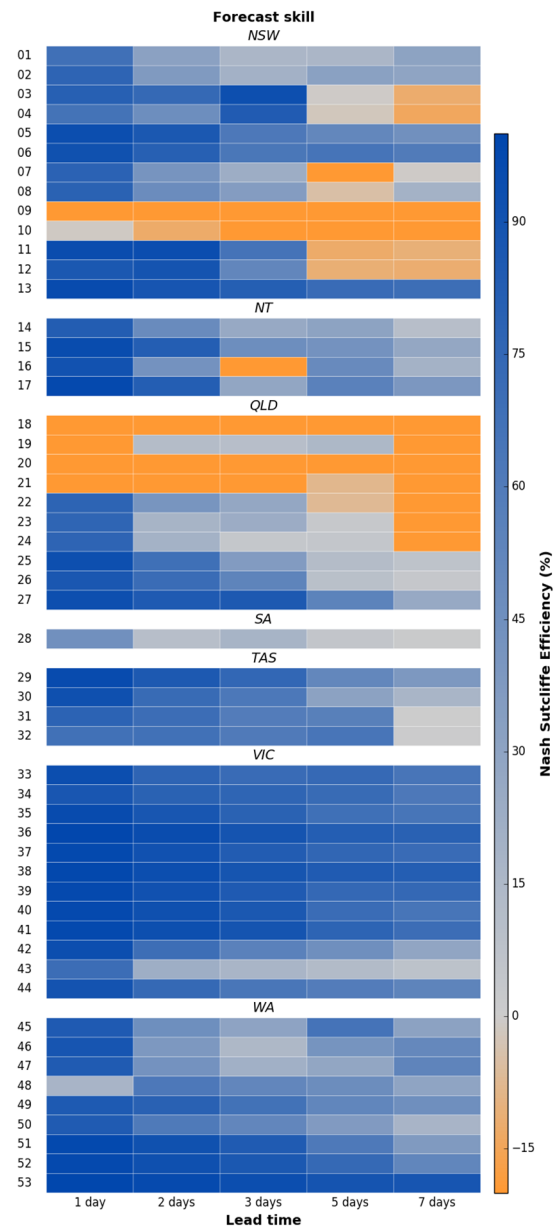


Figure 3. Streamflow forecast skill at forecast locations with lead-time. The numbers refer to forecast locations given in Table 1.

[website](#)). The main reasons for lower performance for 10% of the forecast locations was driven by erroneous or low quality input data in real time situations, and also by the short (21 month) evaluation data sample size.

Overall, forecast performance for perennial catchments is found to outperform those in the ephemeral catchments. In addition, the forecast performance for locations with large upstream catchment areas (>10000km²) show better skills than those with small catchment areas due to streamflow persistence adding more value than forecast rainfall. Apart from forecast rainfall discrepancies, lack of flexibility of the hydrological model structure to represent complex and highly non-linear hydrological processes in dry, ephemeral catchments is seem to be the cause for suboptimal streamflow forecasts. It is intended to provide a probabilistic streamflow forecast that brings greater confidence and reliability to this service by 2019.

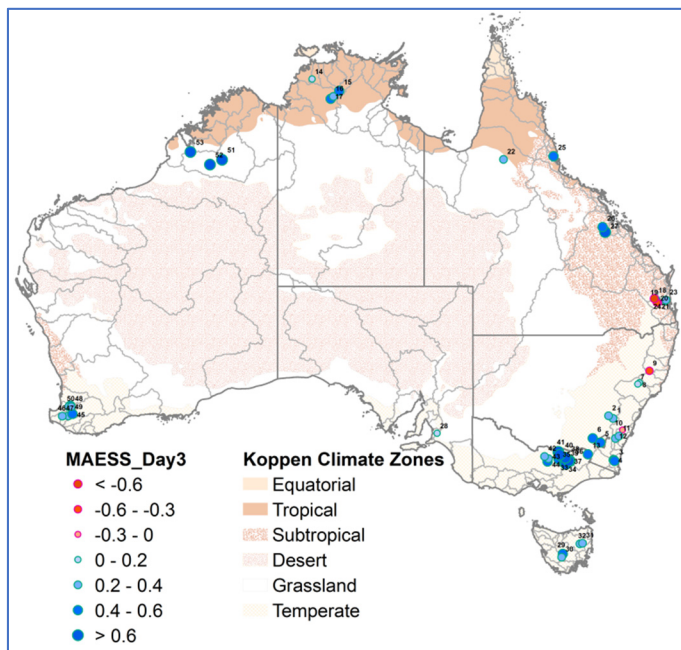


Figure 4. Mean absolute error skill score at 3-day lead-time at selected forecast locations. Numbers next to forecast locations refer to forecast locations given in Table 1.

ACKNOWLEDGMENTS

We thank the Bureau’s HyFS development team, the Flood Forecasting and Warning team, Research and Development group, Information Systems and Services, and the Communications team for their support. Research underpinning this service was conducted by CSIRO through the WIRADA program. We thank all the reviewers for their careful reading of the manuscript and insightful comments and suggestions.

REFERENCES

Atkinson, R., Power, R., Lemon, D., O’Hagan, R., Dovey, D., Kinny, D. (2008). The Australian Hydrological Geospatial Fabric – Development Methodology and Conceptual Architecture CSIRO: *Water for a Healthy Country National Research Flagship*, Canberra, Australia, 57p.

Hapuarachchi, P., Kabir A., Zhang, S., Tuteja, N., Enever, D., Kent, D., Bari, M., Shin, D., (2016). Development of an operational system for the 7-day streamflow forecast service in Australia, *Proceedings of OZWater conference*, 10-12 May, Melbourne.

Cunge, J. A., (1969). On the Subject of a Flood Propagation Computation Method (Muskingum Method), *Journal of Hydraulic Research*, Vol. 7, No.2

Kelly J., Donaldson, A., Ryan, C., Bally, J., Wilson, J., Potts, R. (2004). The Australian Bureau of Meteorology’s next generation forecasting system. *20th International Conference on Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology*.

Pagano, T., Hapuarachchi, P., Shrestha, D., Ward, P., Anticev, J., Wang, Q.J. (2011). Hydrologic modelling with Short-term Water Information Forecasting Tools (SWIFT) to support real-time short-term streamflow forecasting. In WIRADA Water Information Research and Development Alliance: *Science Symposium Proceedings*. pp. 99–105.

Perrin, C., Michel, C., Andréassian, V. (2003). Improvement of a parsimonious model for streamflow simulation. *Journal of Hydrology*, 279: 275-289, doi: 10.1016/S0022- 1694(03)00225-7.

Raupach, M. R., Briggs, P. R., Haverd, V., King, E. A., Paget, M., Trudinger, C. M. (2008). Australian Water Availability Project (AWAP) CSIRO Marine and Atmospheric Research Component: *Final Report for Phase 3*, CSIRO, Canberra, Australia.

Ward P., Robertson D., and Hapuarachchi H. A. P. (2012). The Short-term Water Information Forecasting Tools user Manual, The SWIFT User Manual