

# Estimation of road closure risks along the Bruce highway using the AWRA-L water balance model

E. Vogel<sup>a</sup>, J. Lerat<sup>a</sup>, L. A. Gericke<sup>b</sup>, C. A. Russell<sup>b</sup>, A. Preece<sup>a</sup>, G. Stuart<sup>a</sup>, R. Pipunic<sup>a</sup>,  
Z. Khan<sup>a</sup> and C. Donnelly<sup>a</sup>

<sup>a</sup> Bureau of Meteorology, Melbourne/Canberra/Brisbane, Australia

<sup>b</sup> Queensland Department of Transport and Main Roads, Brisbane, Queensland, Australia  
Email: [elisabeth.vogel@bom.gov.au](mailto:elisabeth.vogel@bom.gov.au)

**Abstract:** The Bureau of Meteorology, in collaboration with the Queensland Department of Transport and Main Roads, conducted a case study that aimed to evaluate the performance of nine-day runoff forecasts in predicting road flooding for selected river crossings along the Bruce Highway. The Bruce Highway is a major traffic carrier in Queensland and one of the most critical transport infrastructures in the state. Flooding in one of the close rivers along the highway can cause the road to be closed. Skilful forecasts of the risk of flooding several days ahead would provide an opportunity to prepare for the event and reduce any negative impacts for commuters on the highway. We developed and tested a method to relate simulated catchment runoff from the national AWRA-L water balance model with water levels at gauging stations close to major highways in Queensland. We found promising potential of predicting the risk of flooding in the next five days, as highlighted by high hit rates (average of 67.9% for moderate flood level at 15 stations) and low false alarm rates (average of 0.5%) for most stations and flood thresholds. This paper provides an overview of the evaluation results and outlines recommendations for future work.

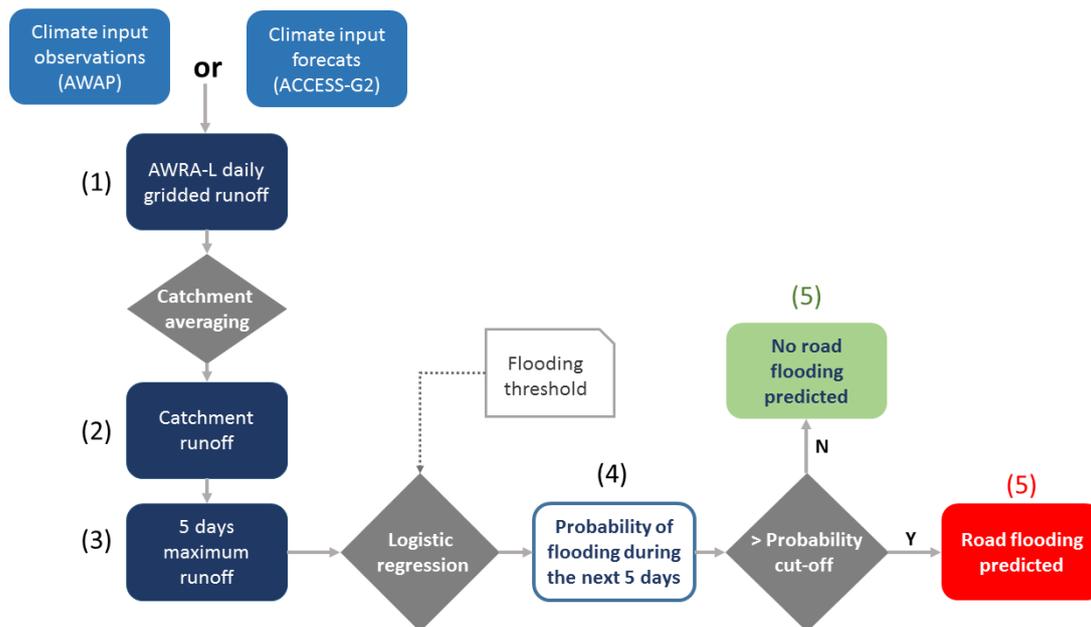


Figure 1. Flow diagram illustrating the generation of a road-flooding forecast.

**Keywords:** Short-term forecasts, flooding risk, transport infrastructure

## 1. INTRODUCTION

Flooding is a recurrent natural hazard that has a significant impact on Australian communities and the economy. In Queensland, extreme weather conditions and isolation of inland town centres accentuate the impact of flooding, especially on transport infrastructure. Rolfe *et al.* (2013) estimated that the cost of road closure around Rockhampton during the 2010/2011 floods reached \$3.3 million per day of disruption. The significant impact of road closure due to flooding on transport projects prompted the Queensland Department of Transport and Main Roads (TMR) to include it in its cost-benefit analysis manual (TMR, 2011, section 2.41 and 5.5). More precisely, the guideline stresses that "Flood warning times and the availability of alternative routes will (...) affect the decisions made by road users" (TMR, 2011, page 3.116).

Being responsible for weather and flood warnings across Australia, the Bureau of Meteorology (BoM) is well placed to assist TMR with road flooding forecasts. To explore this avenue, the Bureau and TMR collaborated on a preliminary analysis to quantify the value of runoff forecasts generated with the AWRA landscape water balance model (AWRA-L) for road flooding prediction. The analysis was undertaken as part of the development by the Bureau of a short-term to seasonal landscape forecasting system for hydrological variables – including soil moisture at different depths, runoff and evapotranspiration. Previous assessment of forecast skill has shown good performance for different variables, regions and lead times and promising potential for a wide range of applications.

This paper presents a forecast evaluation focused on the prediction of road flooding for selected river crossings along the Bruce Highway (flooding hotspots). The Bruce Highway is a major highway that connects the state capital, Brisbane, with Cairns in Far North Queensland, passing major coastal centres, including Townsville, Mackay, Rockhampton and Maryborough. Being the biggest traffic carrier in Queensland, it is one of the most critical transport infrastructures in the state. Flooding in one of the close rivers along the highway can cause the road to be closed. Skilful forecasts of the risk of flooding several days ahead would provide an opportunity to prepare for the event and reduce any negative impacts for commuters on the highway. Here, we developed a statistical post-processor based on logistic regression that relates AWRA-L runoff to the risk of road flooding and evaluated predicted road flooding probabilities against historical observed water level data.

## 2. METHODS

### 2.1. Data

The analysis in this study is based on three data sources: (i) water level time series at several gauging stations provided by the state governments and compiled in the Bureau of Meteorology Wiski database; (ii) Road flooding level thresholds provided by the Bureau of Meteorology, and (iii) simulated catchment-runoff using the AWRA-L model. The datasets used in this study generally cover the period 1960-2018 (streamflow and water level data as well as AWRA-L simulations using AWAP climate inputs), with varying levels of data availability between stations.

#### *Site selection*

We selected 15 streamflow and water level gauging stations in Queensland that fulfilled the following criteria: (i) Location within 50 m of major highways in Queensland, (ii) availability of long-term high-quality measurements of streamflow and water level, and (iii) available information on flooding thresholds (i.e. flood levels and bridge height). The locations of the gauging stations are presented in Figure 2.



**Figure 2.** Location of selected gauging stations

#### *Road flooding levels – physical thresholds and percentile-based thresholds*

For 13 out of 15 stations, the flood level database of the Bureau of Meteorology (accessible via: <http://www.bom.gov.au/qld/flood/networks/index.shtml>) contained data on minor, moderate and major flood water levels. However, such information might not always be available, especially in remote locations.

Therefore, we used two types of road flooding thresholds: (i) Physical flooding thresholds, i.e. flood thresholds corresponding to moderate and major flood heights for stations where such data is available; and (ii) Percentile-based flooding thresholds: We found that the minor, moderate and major flood levels averaged to be close to the 99<sup>th</sup>, 99.5<sup>th</sup> and 99.9<sup>th</sup> percentiles, respectively. In other words, the three flood levels correspond to the top 1%, 0.5% and 0.1% of measured water levels at each station. Therefore, we used these as empirical (percentile-based) thresholds for stations where no flood thresholds were available.

## 2.2. Modelling of runoff using AWRA-L

### *AWRA-L model*

The runoff was generated using the AWRA-L water balance model v6 (Frost *et al.*, 2018). AWRA-L is a gridded water balance model that simulates land surface water fluxes and stores, including runoff, evapotranspiration, deep drainage and soil moisture for three soil layers (top: 0-10 cm, shallow: 10-100 cm, deep: 100-600 cm), on a 5km x 5km grid. The model was developed by CSIRO and BoM and is run operationally by the BoM to provide hydrological data from 1911 until present ([www.bom.gov.au/water/landscape](http://www.bom.gov.au/water/landscape)). The AWRA-L model has been calibrated to observed streamflow, catchment-averaged soil moisture and evapotranspiration across Australia and extensively evaluated against in-situ hydrological observations, including streamflow, soil moisture, evapotranspiration and groundwater recharge (Frost and Wright, 2018). It is used to inform the National Water Accounts, water resource assessments, monitoring of soil moisture conditions in Australia and registered users. The input variables of AWRA-L v6 are: daily precipitation, minimum and maximum air temperature, wind speed (at 2m) and solar radiation.

### *Computation of catchment runoff from AWRA-L simulations*

AWRA-L gridded runoff was processed in three steps for each forecast location: 1) The boundary of the catchment drained at the location was first delineated using a GIS tool and the Geofabric dataset (<http://www.bom.gov.au/water/geofabric/>). 2) The AWRA-L daily runoffs time series for the cells located within the catchment were extracted and averaged to compute a simulated catchment runoff. 3) Finally, a running 5-days maximum was applied to the catchment runoff daily series. The 5-day maximum was chosen as AWRA-L has no implementation of catchment routing yet (a routing method is currently being developed). Due to this, the timing of maximum runoff in AWRA-L may not always follow the actual flood timing, especially in larger catchments. However, the model may nonetheless be used to flag high-risk periods for road flooding aggregated for multiple days.

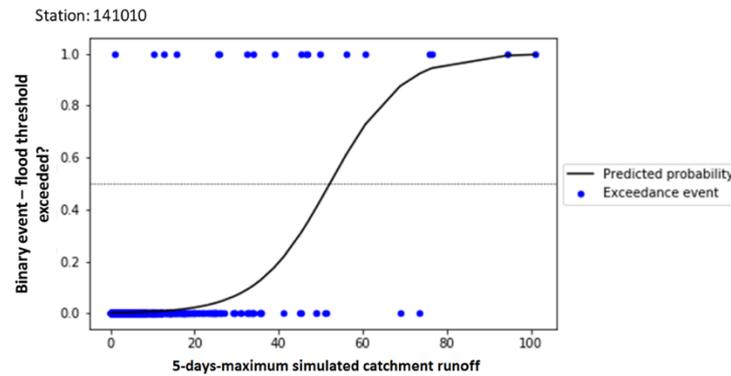
### *Climate inputs used by AWRA-L*

Two types of climate inputs were used as inputs to AWRA-L:

- 1) Climate observations: the maximum predictive capacity of AWRA-L catchment to forecast the risk of flooding was assessed using a historical AWRA-L runoff simulation based on meteorological observation-based datasets: precipitation and temperature from the AWAP dataset (Jones *et al.*, 2009), station-based wind speed data (McVicar *et al.*, 2008), and remote-sensed solar radiation (Grant *et al.*, 2008). All data were available on a 5km x 5km grid.
- 2) Climate hindcasts: retrospective forecasts of catchment runoff (referred to as “hindcasts”) were used to assess the reliability of the nine-day ahead AWRA-L runoff forecasts. Runoff hindcasts were created by running AWRA-L with hindcast data from a numerical weather prediction model (ACCESS-G2) which produces 10-day forecasts of daily rainfall, temperature, wind speed and solar radiation, among other climate variables. A hindcast is produced assuming that climate data is available up to a specific forecast date (in the past) – after this date the model is run in forecast mode, i.e. without any inputs from historical observations. ACCESS-G2 hindcast are available from April 2016 to January 2018.

## 2.3. Statistical post-processing of simulated catchment runoff

A logistic regression was used to identify the relationship between catchment runoff (predictor) and whether or not a road flooding threshold is exceeded (binary predictand). The regression predicted a probability of occurrence using a Sigmoid function as shown in Figure 3. The probability of occurrence is subsequently converted to a binary prediction (flood will occur or not) by comparing the predicted probability with a cut-off level (for example a cut-off of 50% is used in Figure 3 as shown by the dotted horizontal line).



**Figure 3.** Sigmoid function underlying the logistic regression fitted for station 141010.

The training of the logistic regression required the definition of a series of road flooding events from observed water level data. To explore the sensitivity of the post-processor to the threshold, four different flood thresholds were used: 1) Physical-based moderate flood level, 2) Physical-based major flood level, 3) 99.5 Percentile-based flood level (approximately corresponding to moderate), and 4) 99.9 Percentile-based flood level (approximately corresponding to major). The overall process used to generate a road flooding forecast is illustrated in Figure 1.

#### 2.4. Forecast verification

To assess road flooding forecast performance, we calculated: 1) **Hit Rate**: fraction of flooding events (in %) that were correctly predicted, 2) **False Alarm Rate**: non-flooding events (in %) that were incorrectly predicted as events, 3) **ROC score**: the ROC score is an integral of the fraction of Hit Rate and False Alarm Rate over all p thresholds and describes the usability of the forecast more generally. The ROC score ranges between 0.5 and 1.0, with 0.5 indicating a forecast with no skill and 1.0 being a perfect forecast.

All verification metrics were calculated in a cross-validation approach: 80% of data points were used as training data and the remaining 20% as test data. This step was repeated five times such that all data points were used as test data once. The out-of-sample predictions were subsequently combined and used to compute the verification metrics.

### 3. RESULTS

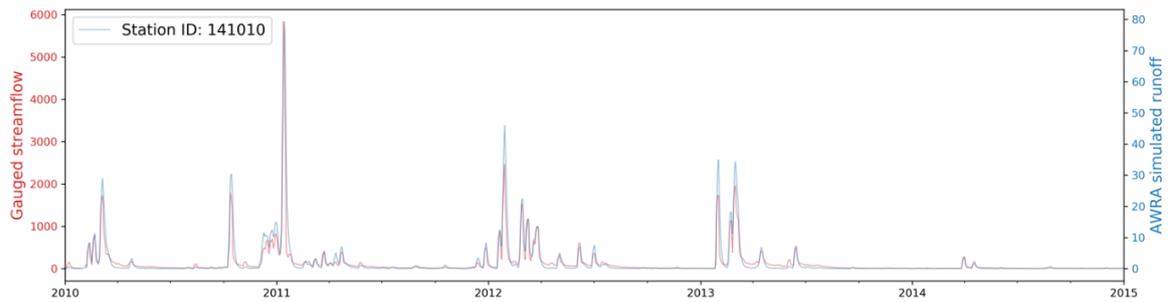
#### 3.1. Results for gauging station 141010 using climate input observations

In this section, we will illustrate the results using one sample location – gauging station 141010 at Coochin Creek, Mawsons Road in the North Coast District, north of Brisbane (see Figure 2 for location). This station was selected due to its location in close proximity to the Bruce Highway and availability of long-term high-quality streamflow and water level data. Results for other stations are reported in subsequent sections.

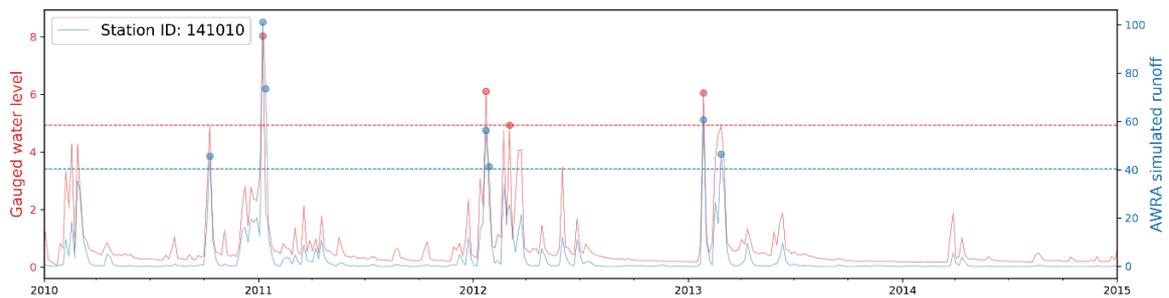
##### *Comparison of catchment runoff with observed streamflow and water level*

Prior to applying the post-processor, we evaluated AWRA-L catchment runoff (step 2 in Figure 1) forced with observed climate against observed streamflow data from the gauge. Figure 4 presents both time series for gauging station 141010. The figure suggests that AWRA-L catchment runoff matches observed streamflow well except for flood peaks and low flow recessions.

Figure 5 shows the time series of observed 5-day maximum water levels (in m) and 5-day maximum catchment runoff (step 3 in Figure 1) for the same station. The dashed line represents the moderate flood level (4.8m) for this station. At this station, ~12.4 years of gauging station data from 2006 to 2018 are available, and the flooding threshold level is exceeded 11 times (1.31% of 5-day periods). Visual examination of the peaks suggests that there is good agreement between peaks in water level and peaks in AWRA-L catchment runoff. Quantitatively, we find that in 73.8% of cases where the maximum water level is in the top 5% of values, the maximum runoff is also in the top 5%. These results suggest that there is promising potential for using AWRA-L runoff to predict critical water level exceedances. This provides the basis for the logistic regression approach used in the following, as we can assume a relationship between simulated AWRA-L runoff and observed water level threshold crossings.



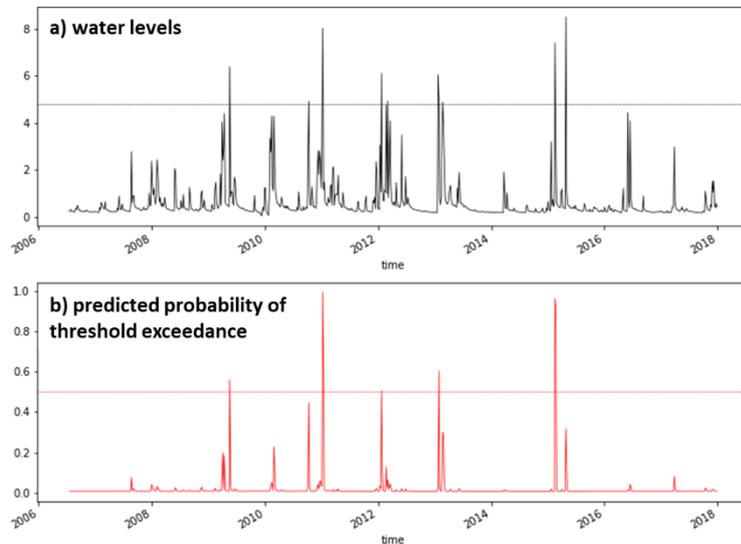
**Figure 4.** Time series of observed streamflow (red, in ML/day) and simulated catchment mean runoff from AWRA-L (blue, in mm/day) for 2010-2015.



**Figure 5.** Time series of 5-day maximum water level (red, in m) and 5-day maximum catchment runoff simulated by AWRA-L (blue, in mm/day) for 2010-2015.

**Probabilistic road flooding forecasts**

Using the logistic regression model, it is possible to visualise the predicted probabilities of a flood event occurring. Figure 5 shows the time series of observed water levels and the timeseries of the predicted probability of exceeding the moderate flood level based on the logistic regression (step 4 in Figure 1). The peaks show good agreement, indicating that the logistic regression is able to capture flood risk events in the historical time series.



**Figure 6.** Time series of observed water levels (a) and timeseries of the predicted probability of exceeding the moderate flood level based on the logistic regression (b), for station 141010.

**Forecast verification**

The Hit Rate and False Alarm Rate depend on the probability cut-off ( $p$ ) that is selected to define an event. By default, a cut-off  $p=50\%$  is selected, i.e. if the forecast probability of an event is greater than 50%, an event is predicted; if the forecast probability is less than 50%, no event is predicted (for example, the dashed line in Figure 5b represents the 50%-threshold). However, other cut-off points are possible and depend on the

decision-making context, such as the societal impacts and costs of a missed event or the costs and effects of false alarms. The Hit Rate and False Alarm Rates for station 141010 for the moderate flood level for three different p-cut off points (10%, 50%, and 66%) are presented in Table 1.

**Table 1.** Hit Rates (HR), False Alarm Rates (FAR) and ROC-area for station 141010 for the moderate flood levels and three probability cut-off thresholds.

		Probability cut-off						ROC-area
		p = 10%		p = 50%		p = 66%		
Flooding threshold	Occurrence rate	HR	FAR	HR	FAR	HR	FAR	
Moderate flood	1.3%	81.8%	1.2%	54.5%	0.1%	54.5%	0.1%	0.927
99.5 pctl flood	0.7%	50.0%	1.1%	16.7%	0.2%	16.7%	0.1%	0.856

There is a trade-off between the Hit Rate and False Alarm Rate – a conservative approach that aims to capture as many events as possible might mean reducing the probability cut-off point. This leads to an increased Hit Rate (more events are predicted), but also generally a rise in the False Alarm Rate. At the same time, if the costs of False Alarms are high and the costs of a missed event are comparatively low, a less conservative approach could mean having a stricter (higher) probability threshold above which an event is predicted. This leads to a lower Hit Rate, but also a lower False Alarm Rate. The balance between Hit Rate and False Alarm Rate – in relation to the p-cut off point – can be re-presented in a plot type referred to as the Receiving Operating Characteristic (ROC) curve (Bradley, 1997). The area under the ROC curve (ROC-area) is a metric for how well the probabilistic forecast distinguishes between events and non-events. The ROC-area ranges between 0.5 and 1.0, with 0.5 indicating a forecast with no skill and 1.0 being a perfect forecast. The ROC-area values of the logistic regression all range above 0.8 – indicating a very high resolution of the prediction model for the out-of-sample predictions (Table 1, see Table 2 for results for the other stations).

### 3.2. Forecast verification for all stations using observed climate inputs

Table 2 generalises the results reported in section 3.1 to all stations. All forecasts assessed here were generated using climate input observations, 99.5 percentile flood threshold (approximately moderate floods) and two probability cut-offs (10% and 50%). The table indicates that false alarm rates remain very low for both probability cut-off, suggesting that the system does not trigger unrealistic flood events. The area under the ROC curve remains also high, revealing that there is no strong trade-off between hit rates and false alarm when changing the probability cut-off. The hit rates vary between 33% and 100% when the probability cut-off is set to 10%, and between 8% and 83% when it is set to 50%. Considering that the false alarms remain low for both cut-off points, it is suggested to use a cut-off of 10%, which delivers much higher hit rates.

## 4. CONCLUSIONS AND RECOMMENDATIONS

We developed and tested a method to generate road flooding forecasts at gauging stations close to major highways in Queensland. We found promising potential of predicting the risk of flooding in the next five days, as highlighted by high hit rates and ROC areas for most stations and flood thresholds. Future work may address several limitations of this study, which are outlined below:

- **Aggregation period:** Due to routing not yet being implemented into the AWRA-L model, we selected a 5-day aggregation period for the analysis, i.e. predicting the risk of flooding within the next five days using the maximum runoff over the same period. Shorter periods were tested as well and show potential of reducing the aggregation period. Future work may investigate the influence of the temporal aggregation period on the verification metrics to select the optimal aggregation period. Furthermore, a routing method is currently being developed in the AWRA-L model, so that a temporal aggregation of runoff may not be required using future model versions.
- **Selection of gauging stations:** In this study, we selected 15 gauging stations based on proximity to major highways and data availability criteria to develop and test the methodology on a small subset of data. In future work, the approach developed may be expanded to the Queensland region more broadly to capture all catchments relevant for TMR operations.

- **Road flooding levels:** The results in this study are based on Road flooding levels in the Bureau of Meteorology river height station network. Future work may use more localised information on critical water levels that have been found to be linked to impacts on the Queensland highway network, reflecting the experiences and impact assessments of TMR.

**Table 2.** Hit Rate and False Alarm Rate for 99.5 percentile-based flood level (approximately moderate flood level) and 15 stations.

Station ID	Number of time steps	Number of events	Ratio: N event / N total	Probability cut-off p = 10%		Probability cut-off p = 50%		ROC
				HR	FAR	HR	FAR	
109001	2554	13	0.51%	69.23%	0.51%	30.77%	0.08%	0.995
116013	3363	15	0.45%	60.00%	0.27%	53.33%	0.06%	0.997
117002	2355	13	0.55%	76.92%	0.64%	7.69%	0.13%	0.993
118106	2586	19	0.73%	52.63%	0.78%	26.32%	0.19%	0.980
120307	1906	12	0.63%	50.00%	0.90%	16.67%	0.00%	0.801
121004	996	6	0.60%	50.00%	0.71%	16.67%	0.00%	0.884
130413	2694	14	0.52%	78.57%	0.45%	50.00%	0.07%	0.996
136209	2146	9	0.42%	100.00%	0.23%	55.56%	0.09%	0.999
137201	3500	19	0.54%	63.16%	0.34%	52.63%	0.14%	0.995
141010	837	6	0.72%	50.00%	1.08%	16.67%	0.24%	0.856
143229	1435	5	0.35%	100.00%	0.14%	60.00%	0.07%	0.999
143901	1156	5	0.43%	80.00%	0.61%	20.00%	0.09%	0.997
416317	1638	9	0.55%	66.67%	0.06%	44.44%	0.06%	0.997
416319	887	6	0.68%	33.33%	0.34%	16.67%	0.11%	0.996
416415	2222	12	0.54%	91.67%	0.27%	83.33%	0.05%	0.999
919013	2806	14	0.50%	64.29%	0.64%	14.29%	0.14%	0.988
Average	2067.6	11.1	0.55%	67.90%	0.50%	35.32%	0.10%	0.967
Min	837.0	5.0	0.35%	33.33%	0.06%	7.69%	0.00%	0.801
Max	3500.0	19.0	0.73%	100.00%	1.08%	83.33%	0.24%	0.999

**REFERENCES**

Bradley, A.P., 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognit.* 30, 1145–1159. [https://doi.org/10.1016/S0031-3203\(96\)00142-2](https://doi.org/10.1016/S0031-3203(96)00142-2)

Frost, A.J., Ramchurn, A., Smith, A., 2018. The Australian Landscape Water Balance model (AWRA-L v6). Technical Description of the Australian Water Resources Assessment Landscape model version 6.

Frost, A.J., Wright, D.P., 2018. Evaluation of the Australian Landscape Water Balance model: AWRA-L v6. A comparison of AWRA-L v6 against Observed Hydrological Data and Peer Models.

Grant, I., Jones, D., Wang, W., Fawcett, R., Barratt, D., 2008. Meteorological and Remotely Sensed Datasets for Hydrological Modelling: A Contribution to the Australian Water Availability Project. *Catchment-scale Hydrol. Model. Data Assim. Int. Work. Hydrol. Predict. Model. Obs. Data Assim.* 1–4.

Jones, D.A., Wang, W., Fawcett, R., 2009. High-quality spatial climate data-sets for Australia. *Aust. Meteorol. Oceanogr. J.* 58, 233.

McVicar, T.R., Van Niel, T.G., Li, L.T., Roderick, M.L., Rayner, D.P., Ricciardulli, L., Donohue, R.J., 2008. Wind speed climatology and trends for Australia, 1975–2006: Capturing the stilling phenomenon and comparison with near-surface reanalysis output. *Geophys. Res. Lett.* 35, L20403. <https://doi.org/10.1029/2008GL035627>

Rolfe, J., Kinnear, S., Gowen, R., 2013. Simplified Assessment of the Regional Economic Impacts of Interruption To Transport Corridors With Application To the 2011 Queensland Floods. *Australas. J. Reg. Stud.* 19, 215–238.

TMR, 2011. Cost-benefit Analysis Manual. State of Queensland, Department of Transport and Main Roads.