Australia's National Science Agency

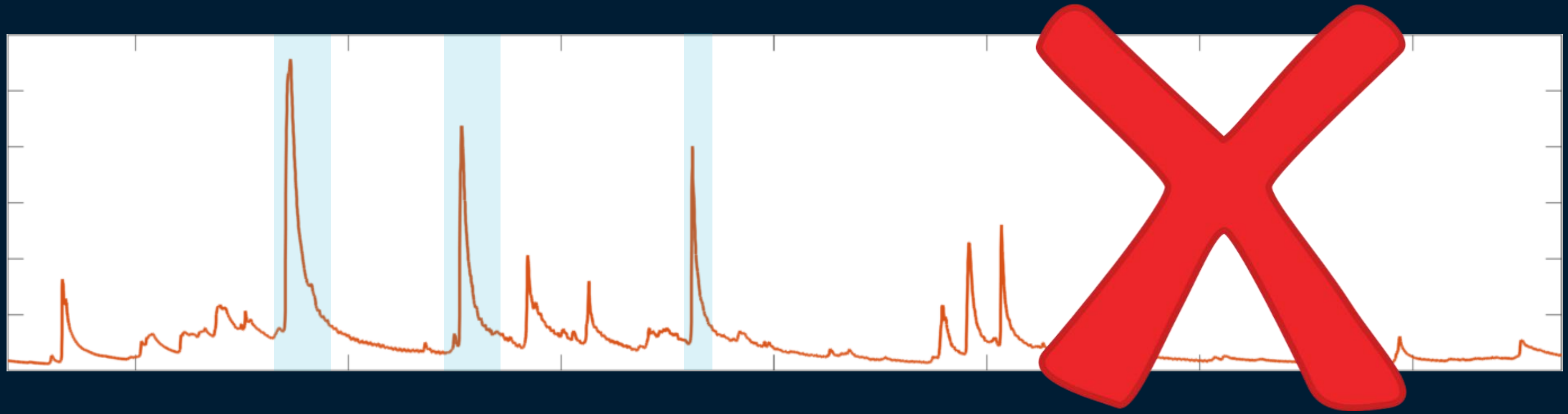
How can we check the reliability of ensemble flood forecasts?

James Bennett & David Robertson | 19 Nov 2020

The problem

- Low streamflows are not like floods!
- How to verify the bits we want?

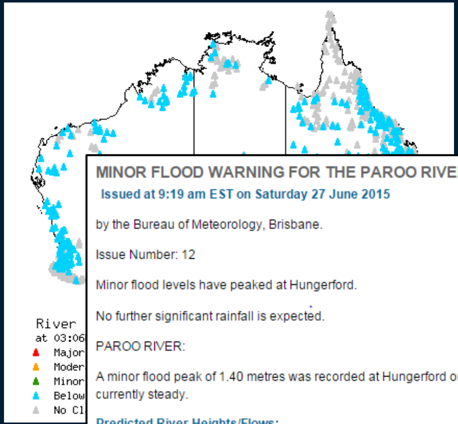
The problem



- Low streamflows are not like floods!
- How to verify the bits we want?

Some context: Aus forecasting services

Flood forecasting



MINOR FLOOD WARNING FOR THE PAROO RIVER

Issued at 9:19 am EST on Saturday 27 June 2015

by the Bureau of Meteorology, Brisbane.

Issue Number: 12

Minor flood levels have peaked at Hungerford.

No further significant rainfall is expected.

PAROO RIVER:

A minor flood peak of 1.40 metres was recorded at Hungerford on Friday morning. River levels are currently steady.

Predicted River Heights/Flows:

PAROO RIVER at:

HUNGERFORD: River levels will remain steady during Saturday and start to fall on Sunday. River levels may fall below the minor flood level during Monday.

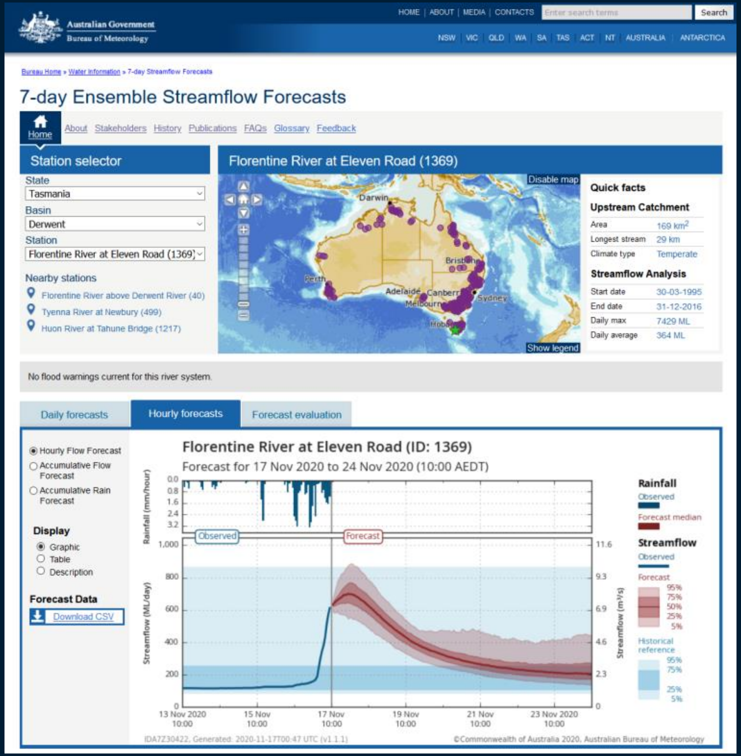
Remember: If it's flooded, forget it.

For flood emergency assistance contact the SES on 132 500

- Flood watch … **situation”**
- Flood warning: **“flooding is occurring or expected to occur”**

http://www.bom.gov.au/water/

7-Day Ensemble service



CSIRO

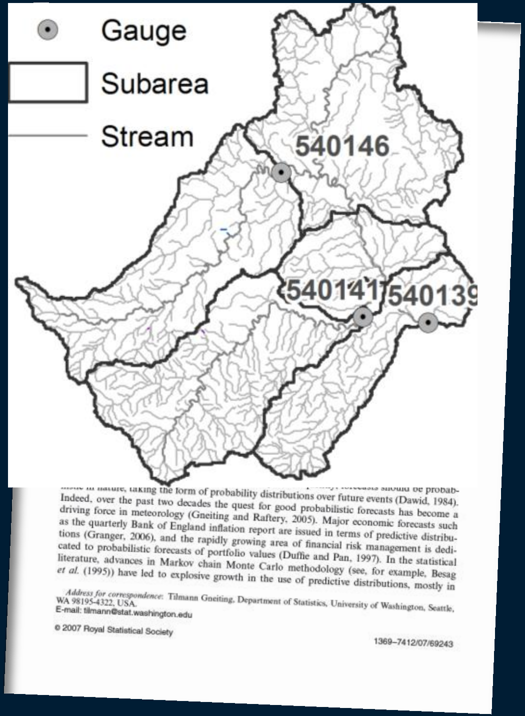
Why care about reliability?

Practical

- Forecast probabilities reflect flood risk
- Uncertainty can be propagated downstream
- Outputs can be used directly
  in decision models

Theoretical

- “Maximise sharpness subject to reliability”
  Gneiting et al.
- No hedging needed



Reliability and forecast selection

- Overall reliability does not guarantee conditional reliability (Hamill 2001)

‘Observation-based’ selection

- Bellier et al. -> Can’t measure reliability
- Lerch et al. -> CRPS not strictly proper

‘Forecast-based’ selection

- Reliability is measurable
- CRPS strictly proper
- False alarms measurable (!)

September 2017 BELLIER ET AL. 3529

Sample Stratification in Verification of Ensemble Forecasts of Continuous Scalar Variables: Potential Benefits and Pitfalls

Joseph Bellier and Isabella Zin

Guillaume Bontron

ABSTRACT

1. Introduction

Forecaster’s Dilem

Forecast Evaluatio

Sebastian Lerch, Thordis L. Thorarins

550

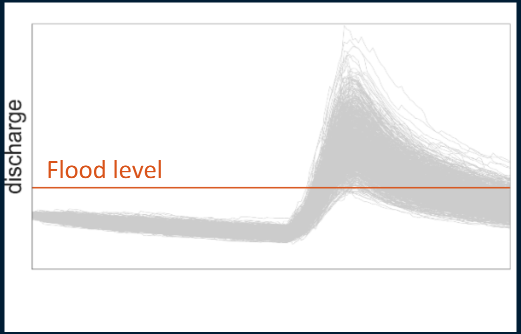
NOTE

Interpretation of Rank

1. Introduction

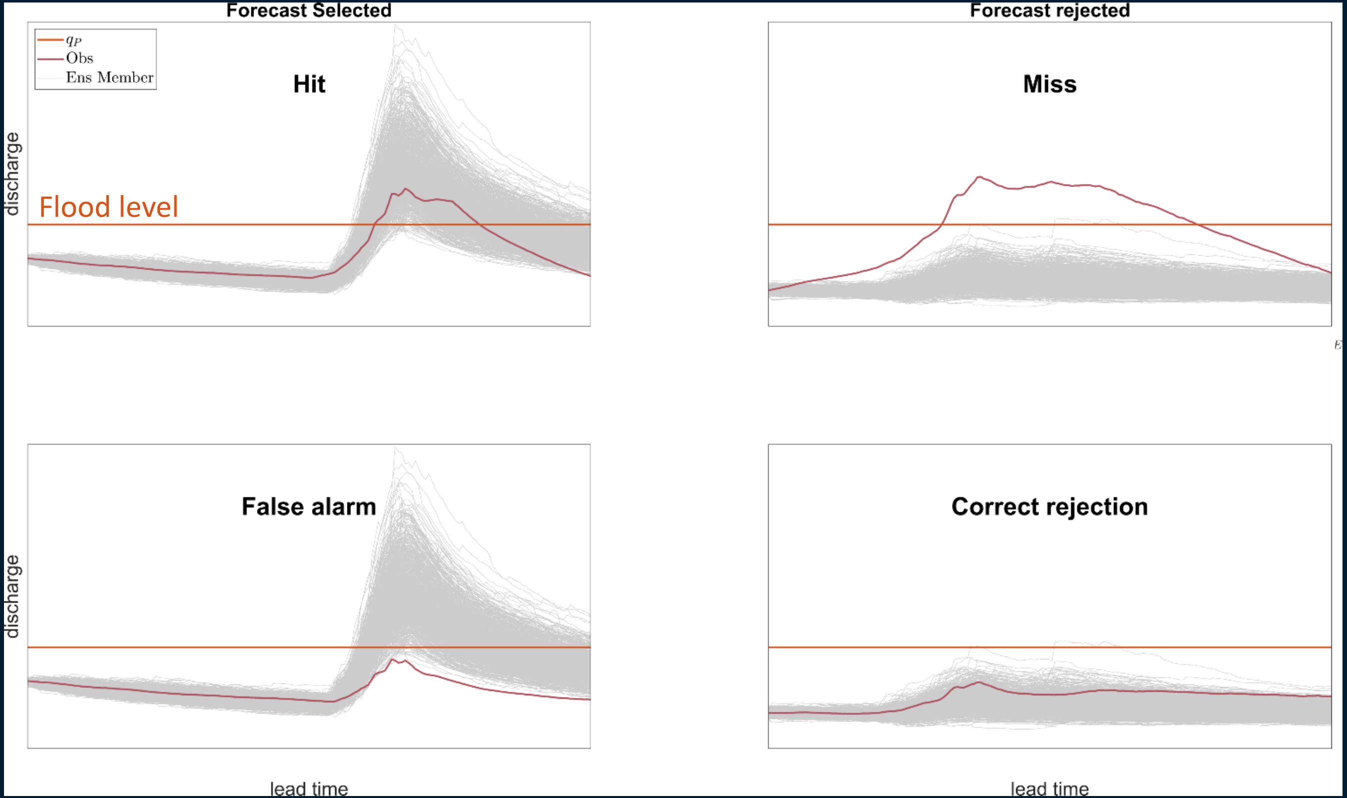
© 2001 American Meteorological Society

Forecast-based selection

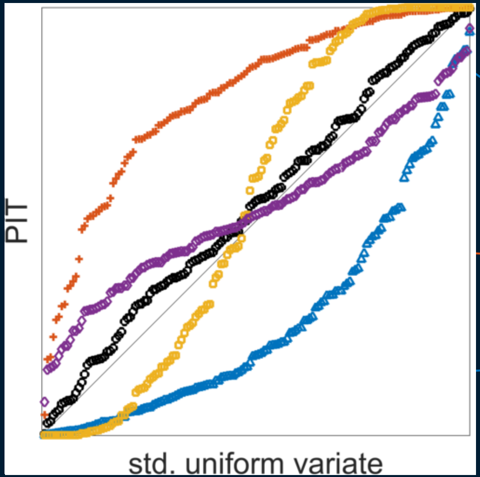


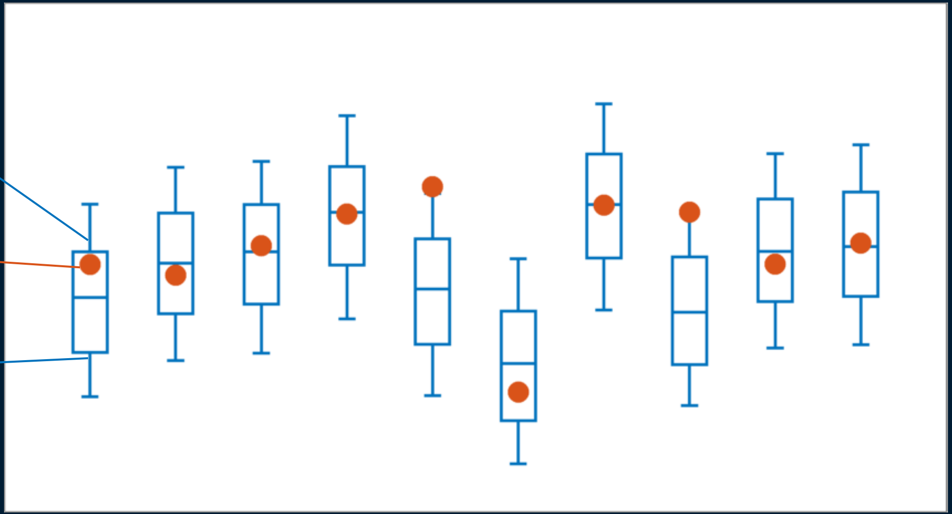
- Heightened probability of flood is forecast -> forecast selected
- Analogous to a ‘flood watch’

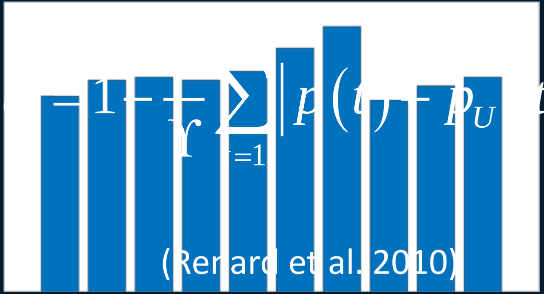
Verifying forecast selection



Reliability – probability integral transform







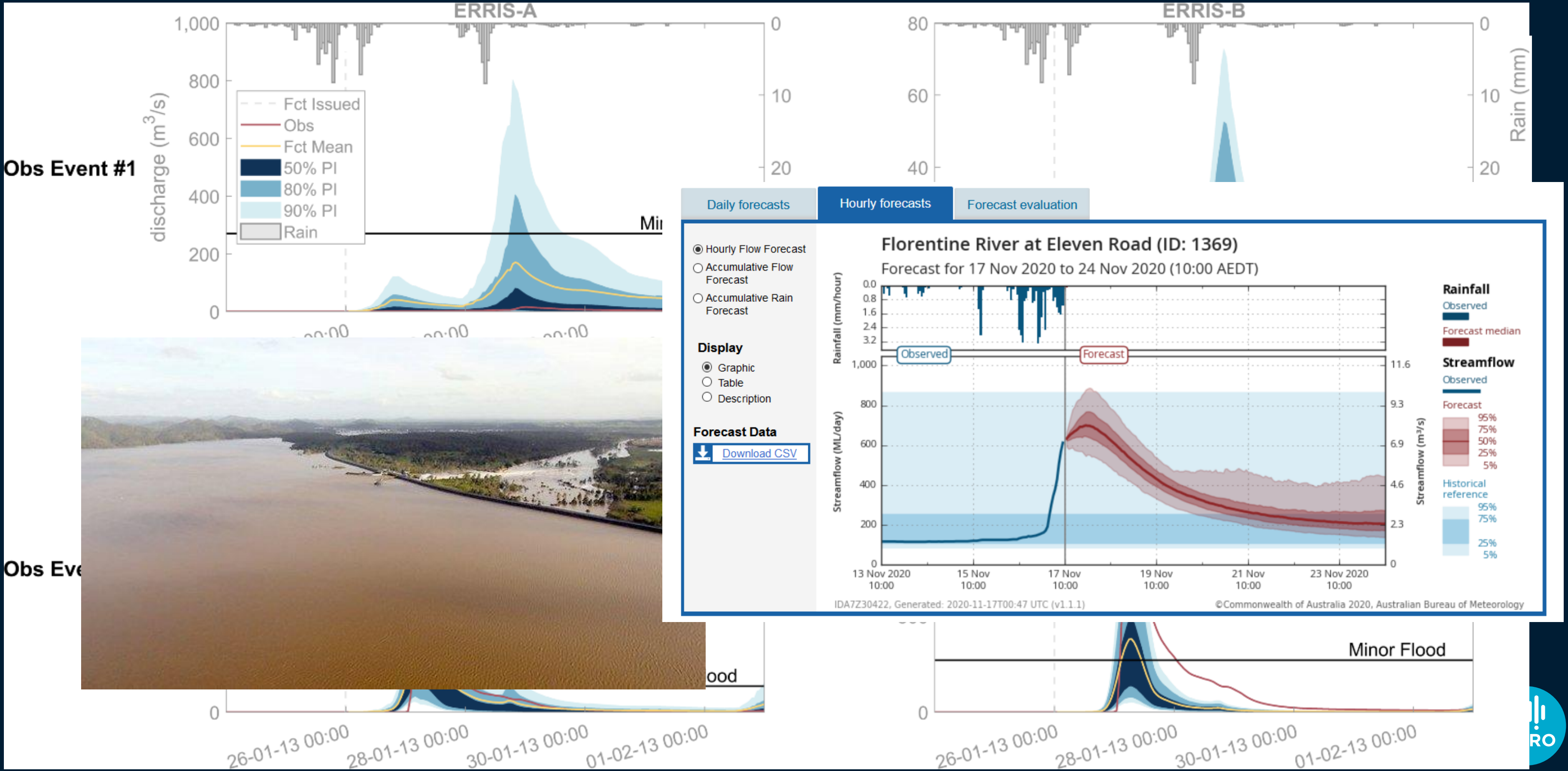
$$= 1 - \frac{1}{Y}\sum_{t=1} \left| p(t) - p_U(t) \right|$$

(Renard et al. 2010)

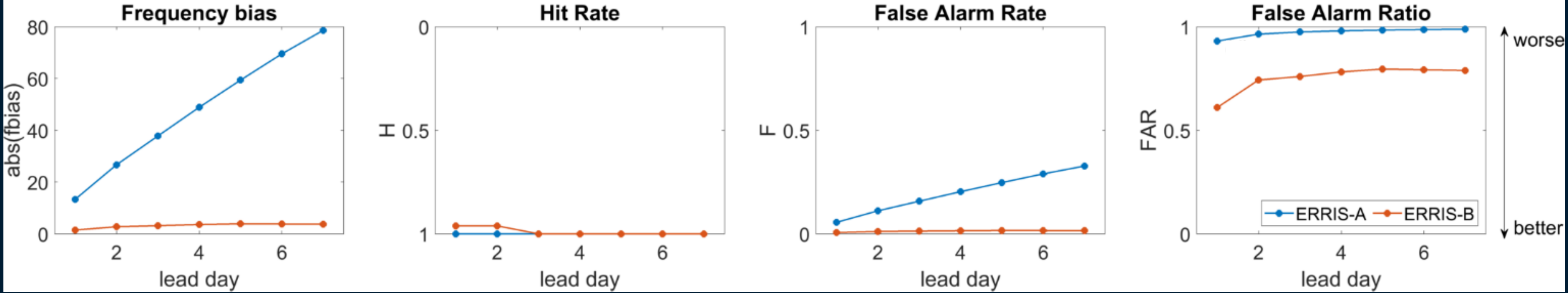
Also need

- Sharpness (AWPI)
- Accuracy (CRPS)

Case study: 2 forecasting methods



Verifying forecast selection



$$\left|\frac{no.\ forecasted\ \ floods - no.\ observed\ \ floods}{no.\ observed\ \ floods}\right|$$

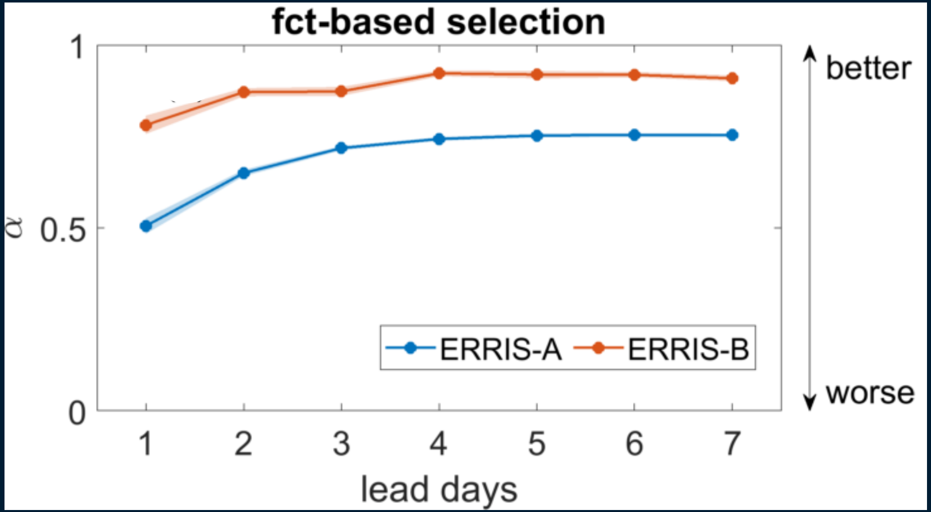
$$\frac{hits}{no.\ observed\ \ floods}$$

$$\frac{false\ \ alarms}{no.\ observed\ \ no\ \ floods}$$

$$\frac{false\ \ alarms}{no.\ forecasted\ \ floods}$$

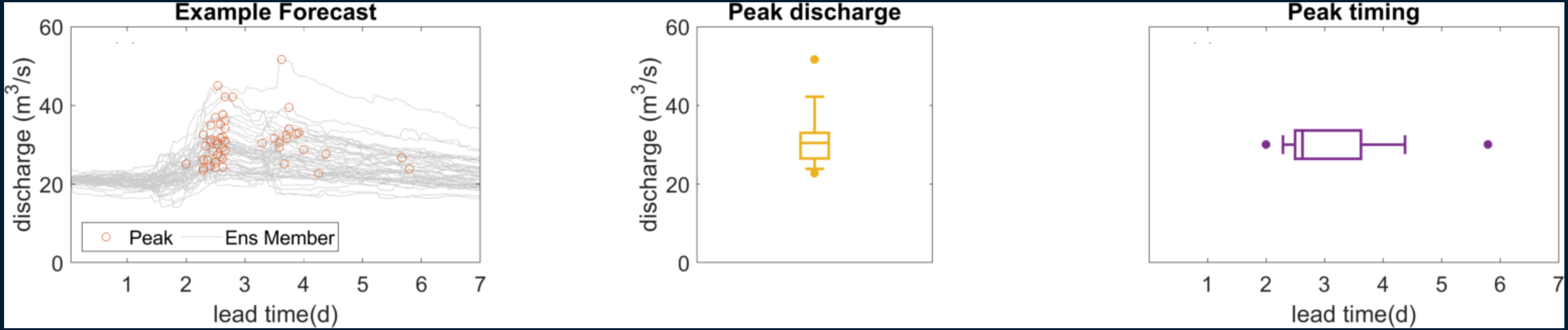
- Flood threshold exceeded in 26 forecasts
- ERRIS-A up to ~2000 forecasts selected
- ERRIS-B up to ~120 forecasts selected

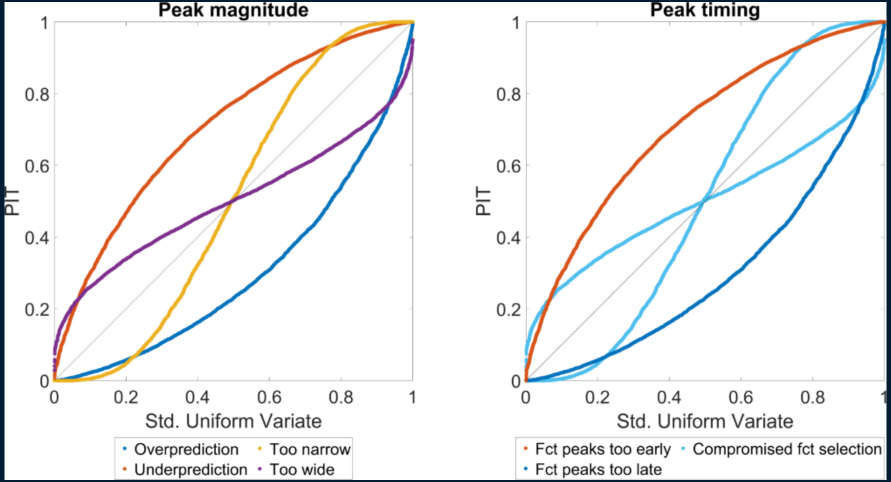
Reliability of flood forecasts



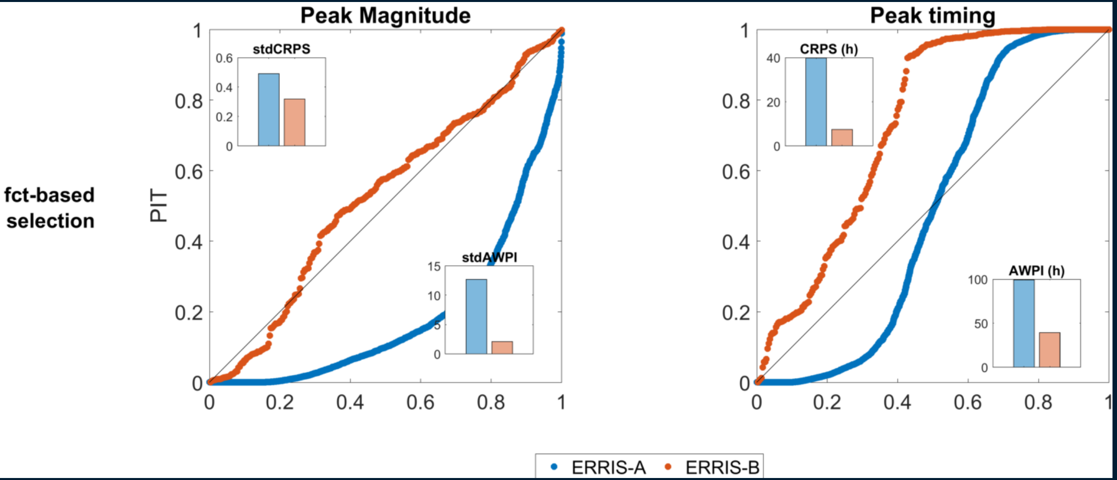
Accuracy and sharpness of flood forecasts

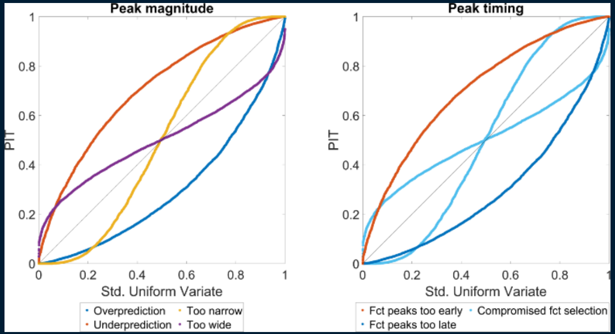
Reliability of flood peak magnitude and timing





Reliability of flood peak magnitude and timing





Recommendations

- Forecast-based selection crucial to measure reliability
  - (and false alarms!)
  - But… forecast-based selection means comparing different numbers of forecasts of different events
- Obs-based selection still useful for communicating performance

References

Forecast verification

Bellier J, Zin I, Bontron G. 2017. Sample Stratification in Verification of Ensemble Forecasts of Continuous Scalar Variables: Potential Benefits and Pitfalls. Monthly Weather Review 145: 3529-3544. DOI: 10.1175/mwr-d-16-0487.1

Gneiting T, Balabdaoui F, Raftery AE. 2007. Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **69:** 243-268. DOI: 10.1111/j.1467-9868.2007.00587.x.

Hamill TM. 2001. Interpretation of Rank Histograms for Verifying Ensemble Forecasts. *Monthly Weather Review* **129:** 550-560. DOI: 10.1175/1520-0493(2001)129<0550:Iorhfv>2.0.Co;2.

Lerch S, Thorarinsdottir TL, Ravazzolo F, Gneiting T. 2017. Forecaster's Dilemma: Extreme Events and Forecast Evaluation. Statist. Sci. 32: 106-127. DOI: 10.1214/16-sts588.

7-day ensemble streamflow forecasting service

Bennett JC, Robertson DE, Shrestha DL, Wang QJ, Enever D, Hapuarachchi P, Tuteja NK. 2014. A system for continuous hydrological ensemble forecasting (SCHEF) to lead times of 9 days. *Journal of Hydrology* **519:** 2832-2846. DOI: 10.1016/j.jhydrol.2014.08.010.

Bennett JC, Robertson DE, Wang QJ, Li M, Perraud J-M. (In review). Propagating reliable estimates of hydrological forecast uncertainty to many lead times. Submitted to Journal of Hydrology.

Li M, Robertson DE, Wang QJ, Bennett JC, Perraud J-M. 2020. Reliable hourly streamflow forecasting with emphasis on ephemeral rivers. Journal of Hydrology 125739. DOI: 10.1016/j.jhydrol.2020.125739.

Robertson DE, Shrestha DL, Wang QJ. 2013. Post-processing rainfall forecasts from numerical weather prediction models for short-term streamflow forecasting. Hydrology and Earth System Sciences 17: 3587–3603. DOI: 10.5194/hess-17-3587-2013.

Shrestha DL, Robertson DE, Bennett JC, Wang QJ. 2015. Improving precipitation forecasts by generating ensembles through postprocessing. Monthly Weather Review 143: 3642-3663. DOI: 10.1175/mwr-d-14-00329.1.

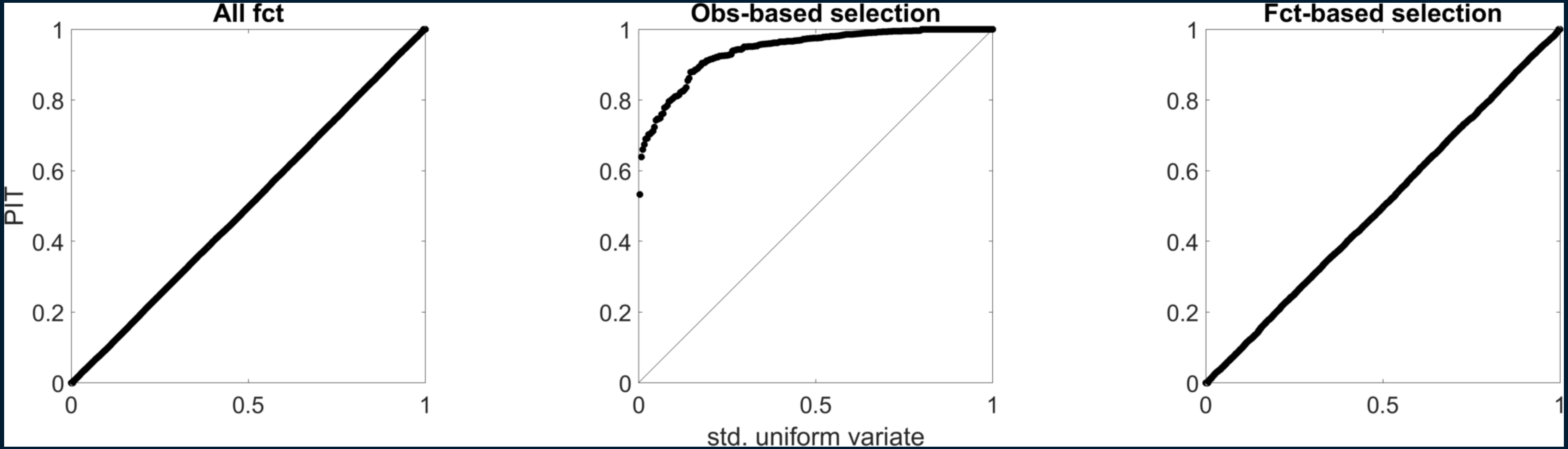
CSIRO

Thank you

Land & Water
James Bennett
Senior Research Scientist

+61 2 9545 2462
james.bennett@csiro.au
https://people.csiro.au/B/J/James-Bennett

Reliability and forecast selection



Synthetic example

- Obs and forecasts drawn from the same normal distributions
- ‘Flood threshold’ based on 99.5% quantile of ‘observations’

Some context: Aus forecasting services