# Advances in improving streamflow predictions, with applications in forecasting

Professor Dmitri Kavetski
Associate Professor Mark Thyer

Intelligent Water Decisions Research Group
(www.waterdecisions.org)
School of Civil, Environmental and Mining Engineering
University of Adelaide

THE UNIVERSITY of ADELAIDE

THE UNIVERSITY OF NEWCASTLE AUSTRALIA

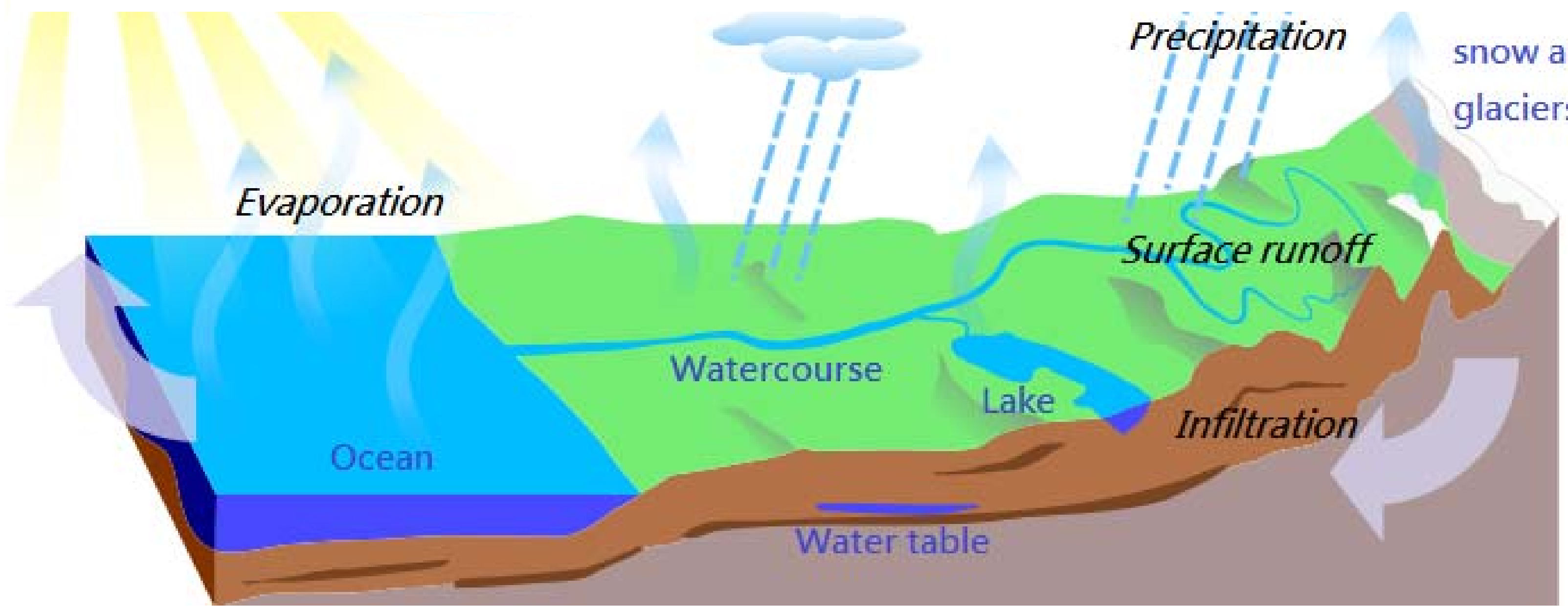
# Acknowledgements

This majority of the outcomes of this presentation represent the combined work of a great team of researchers and operational personnel at the University of Adelaide, University of Newcastle and the Bureau of Meteorology. These include the contributions of Dr David McInerney, Prof. George Kuczera, Dr Narendra Tuteja, Julien Lerat, Daehyok Shin and Fitsum Woldemsekel

The financial support of the Australian Research Council, through the ARC Linkage Grant, LP140100978, the Bureau of Meteorology, South East Queensland Water is gratefully acknowledged.

# Catchment-scale hydrological modelling

- Scientific and engineering hydrology
  - Understanding flow dynamics, improving model “realism”
  - Flood forecasting, drought prediction, and everything in between
- Major component of broader environmental studies
  - Eg, hydrological models provide inputs into larger-scale models such as General Circulation Models (GCMs)




GNU Copyright: Wikimedia: J de Bild and A. Chris

# Practical impact of hydrological models

- Used across many levels of environmental management and decision-making (and increasingly so!)
  - Eg, Australian Bureau of Meteorology streamflow forecasts used for flood forecasting / mitigation, dam operation, water allocations, agricultural uses, etc
- Economic scale of major operations:
  - South East Queensland floods of 2010/11 caused $5 billion damage and impacted 200,000 people (QLD Government, 2013)
  - Australia's irrigated crops: $13.5 billion annual value (ABS, 2013)

=> Credible predictions are expected

=> Must recognize and communicate uncertainties

# Practical impact of hydrological models

- Used across many levels of environmental management and decision-making (and increasingly so!)
  - Eg, Australian Bureau of Meteorology streamflow forecasts used for flood forecasting / mitigation, dam operation, water allocations, agricultural uses, etc
- Economic scale of major operations:
  - South East Queensland floods of 2010/11 caused $5 billion damage and impacted 200,000 people (QLD Government, 2013)
  - Australia's irrigated crops: $13.5 billion annual value (ABS, 2013)

Operational Advances ↔ Research Advances

Even moderate advances = huge value!

# Outline

- Improving Hydrological Predictions
  - Overview of advances in flexible modelling, in lumped and spatial outcomes
  - Improving model calibration / optimization
  - Spatial rainfall modelling
- Improving streamflow probabilistic predictions
  - Importance of persistence for reliable probabilistic streamflow across time scales
  - Practical guidance on residual error models to improve reliability and precision of probabilistic predictions
- Impacts on Forecasting

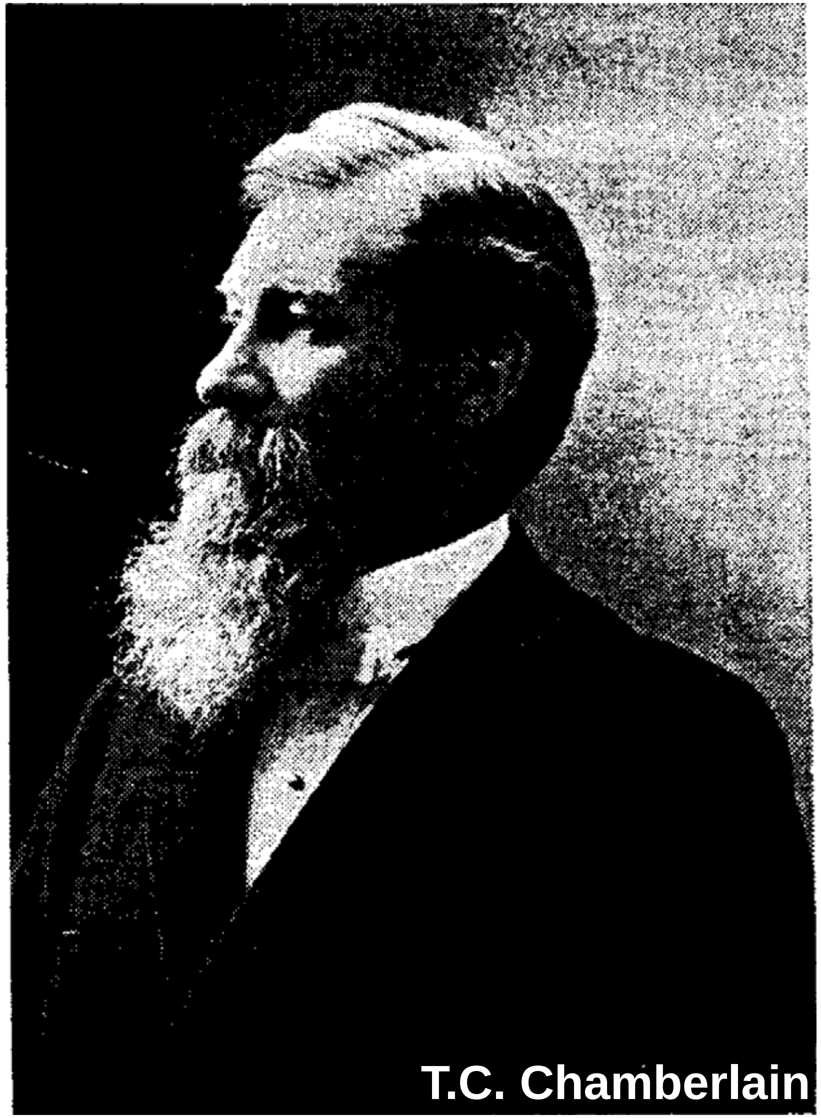
# Flexible models in hydrology: Method of multiple working hypotheses

- Scientists often develop “parental affection” for their theories

- Chamberlin’s method of multiple working hypotheses

*“…the effort is to bring up into view every rational explanation of new phenomena… the investigator then becomes parent of a family of hypotheses: and, by his parental relation to all, he is forbidden to fasten his affections unduly upon any one”*

Chamberlin (1890)



# Multiple working hypotheses in hydrological modelling

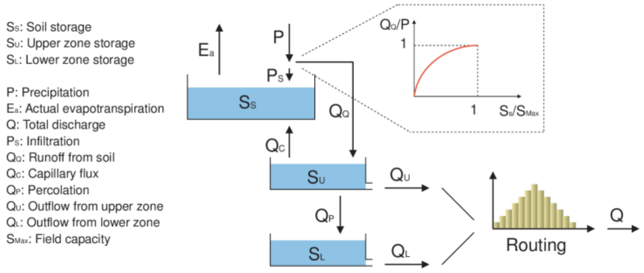
- Multiple alternative rainfall-runoff model structures
  - In simplest case, change numbers of “buckets” in model
  - More generally, include conceptual and physical representations (eg, Fenicia et al 2016; Clark et al 2015)
  - Test using multiple metrics/diagnostics, cross-validation
- Multiple representations of uncertainty
  - Simplest case, eg, unweighted/weighted least squares
  - More generally, different data transformations, etc (eg, McInerney et al, 2017)
  - Again, multiple metrics/diagnostics, cross-validation

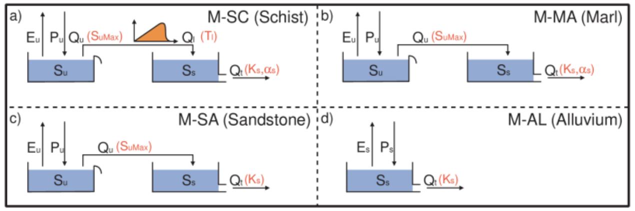
=> More robust findings suitable for operationalization

# Flexible Hydrological Models

## Motivation

- "One-size-fits-all" approaches do not work very well for specific catchment
- Some models, such as GR4J, developed to provide good performance "on average"
- Good starting point, but can we improve performance in catchments of interest?




## Aim

- "Flexible" model that can be tailored to catchments / processes of interest
- Parsimonious => Fast to calibrate/simulate
- Robust => Under cross-validation using multiple metrics (Nash, signatures, etc)

Fenicia F, Kavetski D, Savenije HHG (2010) Elements of a flexible framework for hydrological modelling, *WRR*
Clark et al (2015) A unified approach for process-based hydrologic modeling, *Water Resources Research*
*and references therein / subsequent work*

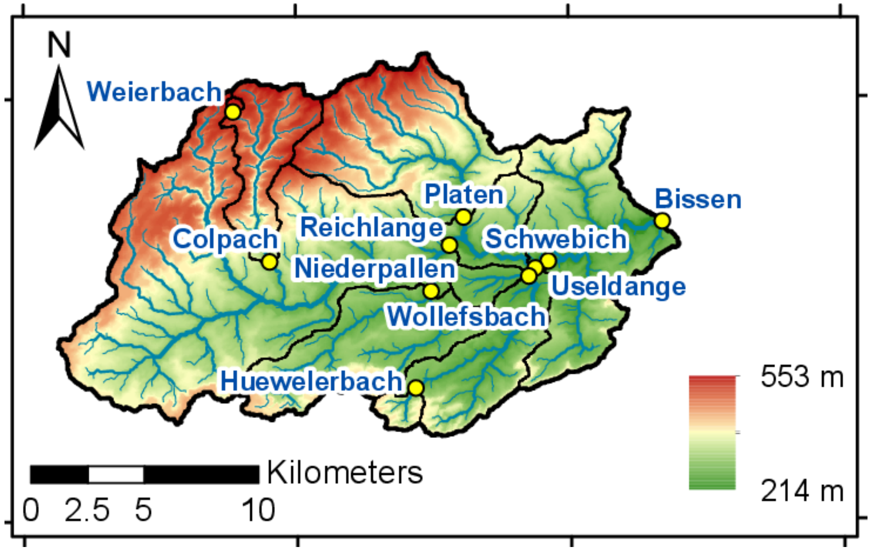
# The challenge of spatial variability

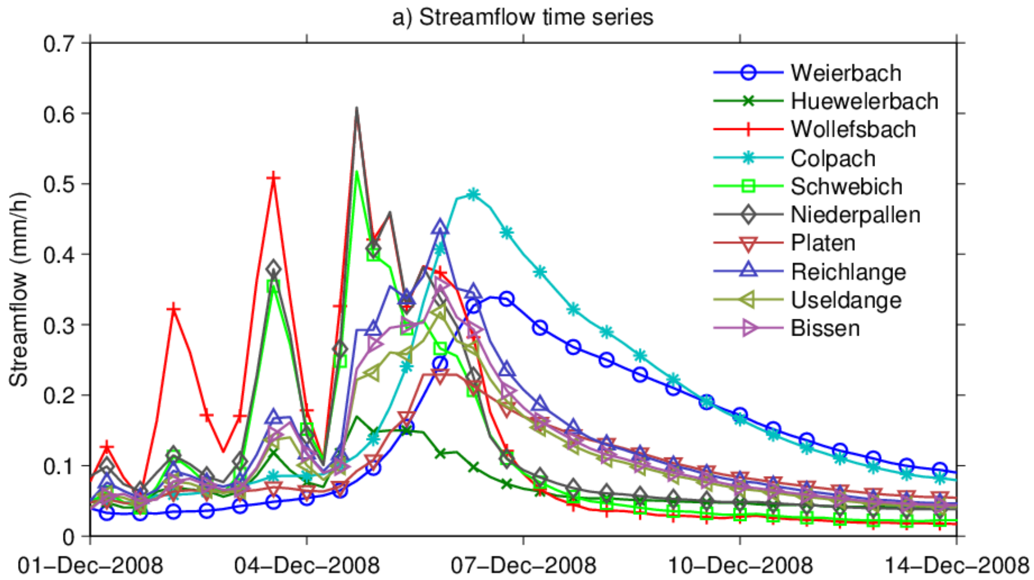
**Study area**

- Attert basin in Luxembourg
- 300 km$^2$
- Densely gauged
- Extensive fieldwork insights

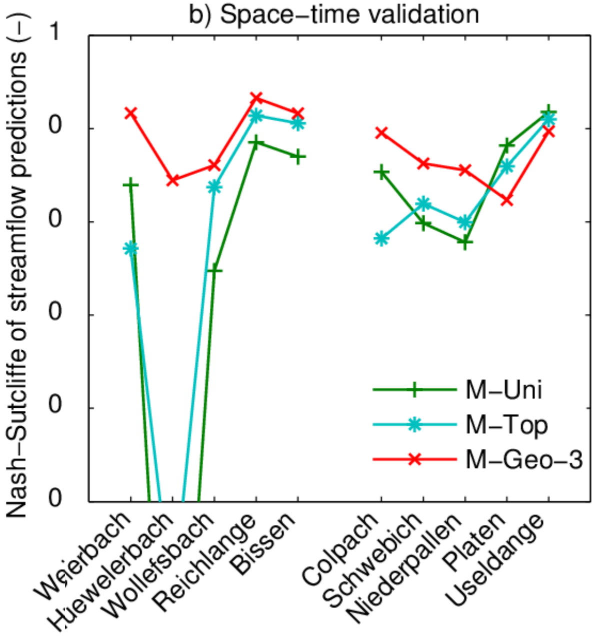
**Research questions**

- Why are the streamflow dynamics so different (dominant controls?)
- How to build a distributed model that reproduces streamflow diversity
- Can such models be exploited for practical / operational prediction?

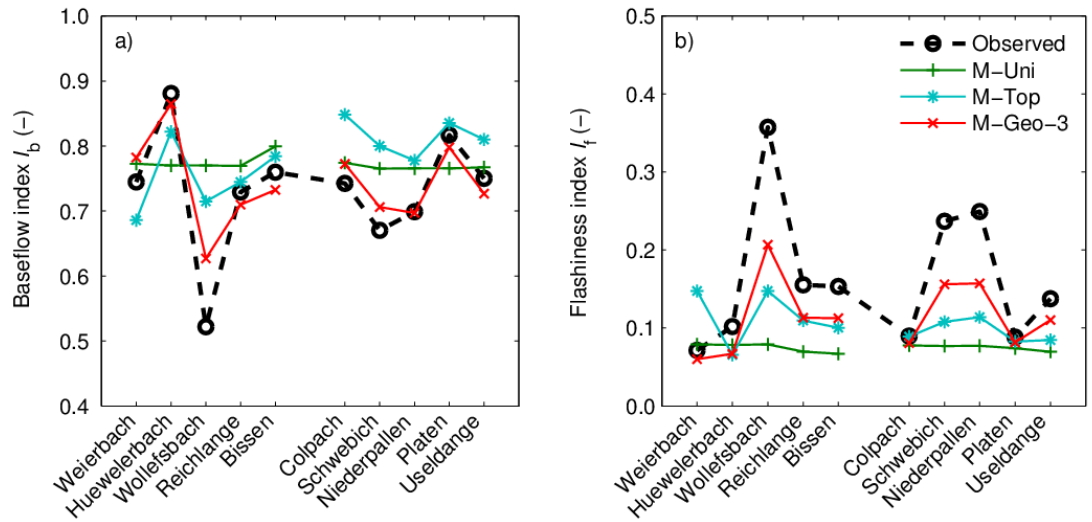


# Distributed application of SUPERFLEX




Use different model structures in each HRU type
Space-time validation, streamflow series and signatures




**Science**: Identify dominant processes, reconcile model vs fieldwork

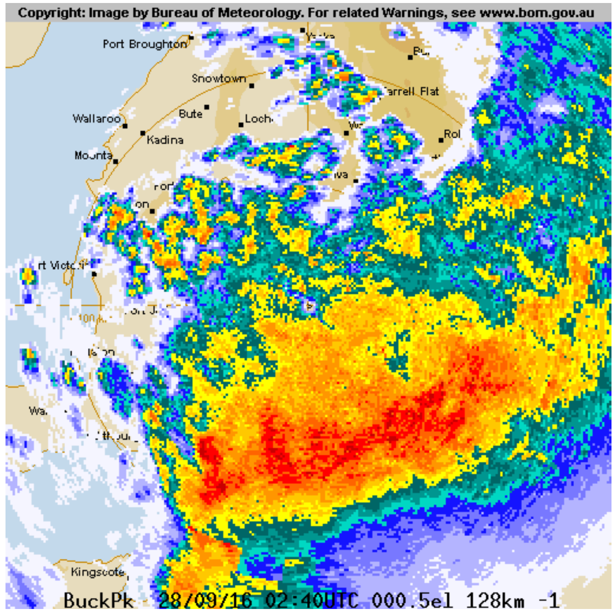
**Operational**: Capture diverse spatial streamflow patterns across 10 locations using “simple” model with 11 calibrated parameters

=> Pilot collaboration with German Forecasting Services

Fenicia et al (2016) From spatially variable streamflow to distributed hydrological models: Analysis of key modeling decisions, *WRR*, 52, 954-989

# Spatial Rainfall Modelling

- Motivation
  - Water management needs stochastic rainfall models to provide reliable estimates of floods / drought risks
  - Spatial rainfall variability has big impact on catchment response
  - Most stochastic spatial rainfall models not simple enough to be used in practice
- Aim
  - Develop stochastic model that continuously simulates daily rainfall fields
  - Parsimonious => Fast to calibrate and simulate
  - Flexible approach => Choose your own spatial resolution

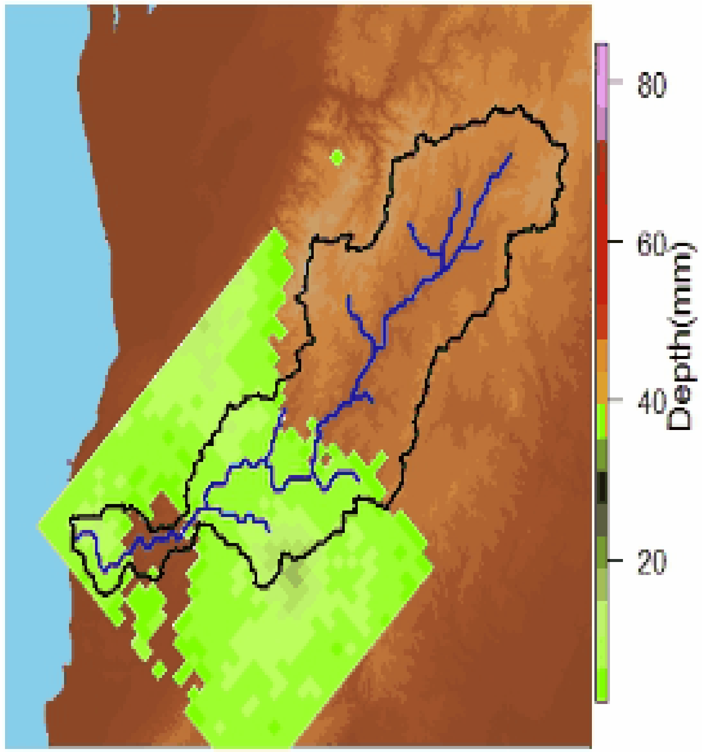



Bennett, B., M. Thyer, M. Leonard, M. Lambert, and B. Bates (2016), A comprehensive and systematic evaluation framework for a parsimonious daily rainfall field model, *J Hydrol* 10.1016/j.jhydrol.2016.12.043

# Realistic continuous rainfall fields

- Daily simulations for Adelaide's water supply catchment: Onkaparinga catchment
- Tested using a comprehensive and systematic evaluation framework
- Majority of spatio-temporal statistics of simulated rainfall were statistically indistinguishable from statistics of observed rainfall
  - Strengths: Rainfall occurrences/amounts, wet/dry spell distributions, Annual volumes/extremes and spatial patterns
  - Weaknesses: Total annual rainfall in dry years (lower 5%) was overestimated.

| | All data | | Cross-validation | |
|---|---|---|---|---|
| Metric | Percent of cases | Overall | Percent of cases | Overall |
| **Monthly** | | | | |
| Wet day amounts – means | 100% | Good | 65% / 19% / 15% | Good |
| Wet day amounts – std dev | 100% | Good | 75% / 18% | Good |
| No. wet days – means | 100% | Good | 64% / 25% / 11% | Good |
| No. wet days – std dev | 44% / 24% / 32% | Highly Variable | 43% / 27% / 30% | Highly Variable |
| Wet spell distribution | 43% / 46% / 11% | Fair – Good | 43% / 46% / 11% | Fair – Good |
| Dry spell distribution | 65% / 35% | Good | 65% / 35% | Good |
| Total rainfall - means | 100% | Good | 44% / 21% / 35% | Highly Variable |
| Total rainfall – std dev | 69% / 28% | Good | 61% / 31% | Good |
| Total rainfall – lower tail | 80% / 18% | Good | 79% / 15% | Good |
| Total rainfall – upper tail | 78% / 20% | Good | 65% / 25% | Good |
| **Annual** | | | | |
| Total rainfall - means | 100% | Good | 11% / 32% / 58% | Poor |
| Total rainfall – std dev | 16% / 47% / 37% | Fair – Poor | 37% / 16% / 47% | Highly Variable |
| Total rainfall – lower tail | 21% / 47% / 32% | Fair – Poor | 26% / 26% / 47% | Fair – Poor |
| Total rainfall – upper tail | 53% / 47% | Good | 53% / 32% / 16% | Good |
| Wet day amounts - mean | 100% | Good | 16% / 42% / 42% | Fair – Poor |
| Wet day amounts – std dev | 100% | Good | 84% / 16% | Good |
| No. wet days – means | 100% | Good | 37% / 42% / 21% | Fair – Good |
| No. wet days – std dev | 100% | Poor | 5% / 95% | Poor |
| **Correlations** | | | | |
| Monthly total rainfall | 89% / 11% | Good | 85% / 15% | Good |
| Annual total rainfall | 100% | Good | 100% | Good |
| **Extremes** | | | | |
| Intensity-Frequency-Duration curve | 53% / 21% / 26% | Good | 37% / 16% / 47% | Highly Variable |



# Potential Applications

- Evaluate impact of spatial rainfall on flood risks
- Evaluate impact of spatial rainfall uncertainty
- Condition stochastic rainfall model on information from numerical weather models
  - Flexible model structure can be adapted to be used for conditional simulation
  - Potential for use as a Spatial Rainfall Post-Processor for rainfall forecast systems (eg, ACCESS)
  - Main advantage is it fills in space, rainfall forecasts over entire grid
  - Opportunity for spatial streamflow forecasts

# Outline

- Improving Hydrological Predictions
  - Overview of advances in flexible modelling, in lumped and spatial outcomes
  - Improving model calibration
  - Spatial Rainfall Modelling
- Improving streamflow probabilistic predictions
  - Practical guidance on which residual error model to improve reliability and precision of probabilistic predictions
- Impacts on Forecasting

# Motivation

- Hydrological predictions are relied upon by wide range of users
- Understanding uncertainty is important for decision making and risk assessment

# Aims

- Overall aim is to improve probabilistic predictions
- Representing uncertainty in hydrological predictions challenging
- We perform comprehensive comparison between approaches for representing uncertainty
- Provide recommendations for practitioners

# Quantifying uncertainty improves decision-making providing better risk estimates

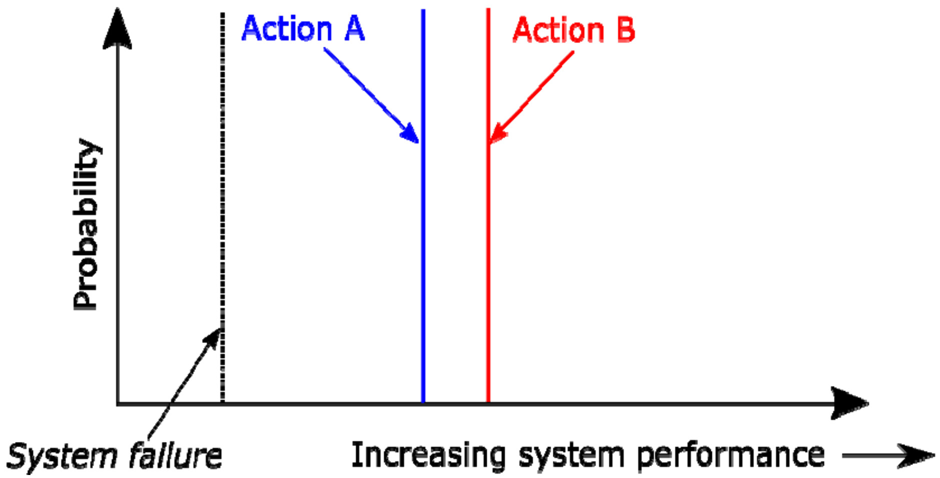
- Example: action A versus action B: Which one would you choose?

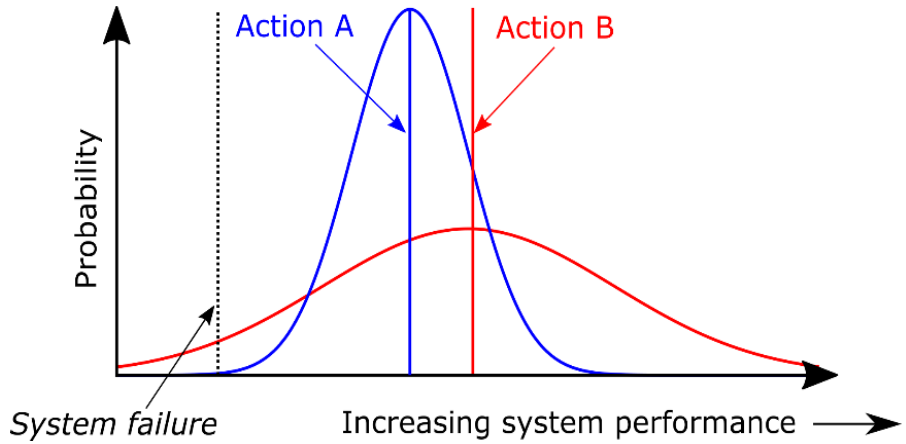
- No uncertainty:

  Use “highest performance” outcome → choose action B!

- With uncertainty:

  If you want to reduce risk of failure→ choose action A!

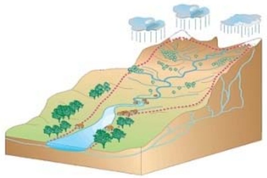


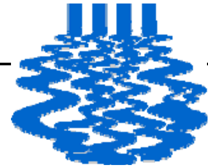




- Water management is all about balancing risks (risk of floods or droughts)
- If we ignore uncertainty, we under-estimates risks of system failure

# Sources of Errors in Hydrological Modelling

Reality





**True Input(s)** → **True processes** → **True response(s)**

**Input errors**

e.g. Rainfall sampling errors

Highly variable in time/space

Low gauge density

**Structural errors**

e.g. Lumping processes

Ubiquitous but

Poorly understood

**Output errors**

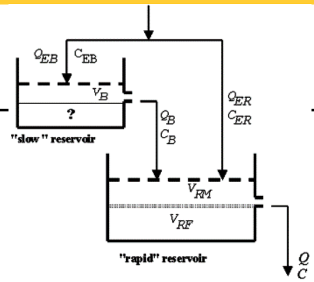
e.g. Rating curve errors

General accuracy ~10-20%

Larger errors in larger floods

**Observed Input(s)** → **Conceptualized processes** → **Observed response(s)**

Model



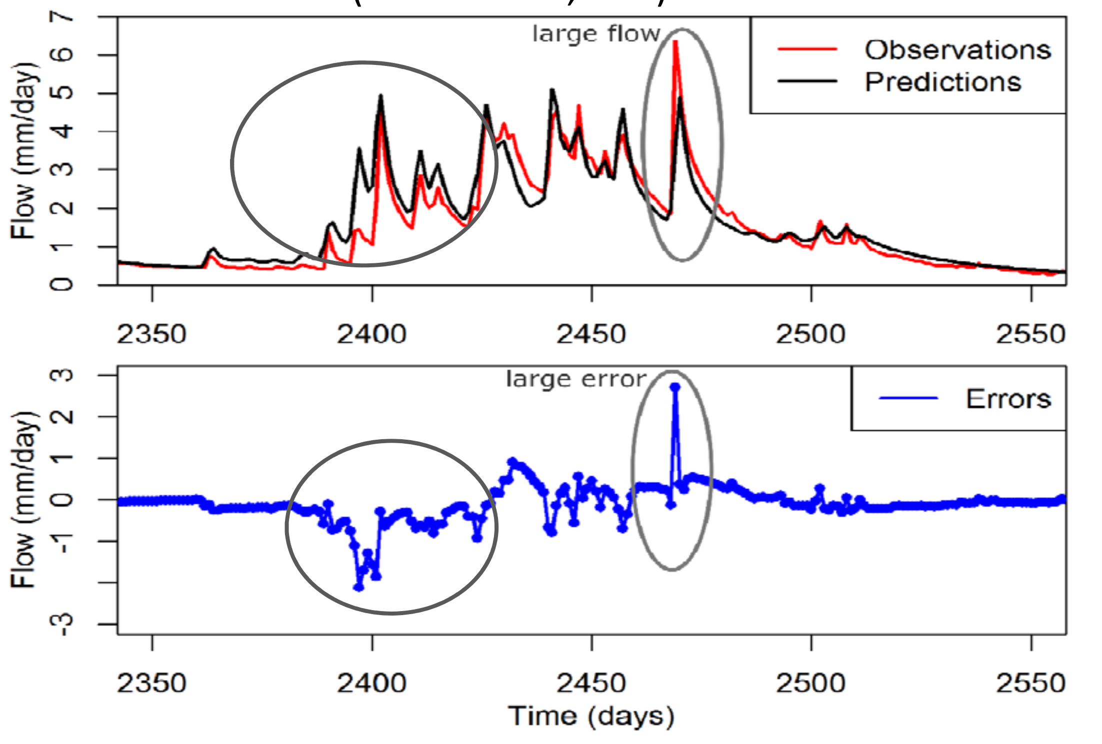

# Approaches to modelling uncertainty: Find the right tool for the job

- Decompositional: Estimate individual sources of uncertainty
  - Advantage of diagnosing dominant sources of error
  - Eg, Bayesian Total Error Analysis (BATEA), Kavetski et al (2006+)
  - Requires more advanced Bayesian and MCMC technology
  - Requires more extensive data support, more expertise in applying, => not really an "off-the-shelf" method available to practitioners
- Aggregated: Estimate total uncertainty in predictions
  - Lump all uncertainty into single term
  - Residual error models: residual = obs – pred, estimate *p*(residual)
  - Easy to articulate, especially in its simpler forms
  - Very common in engineering practice and literature => "off-the-shelf"
  - But: unable to estimate the dominant source of error

In operational settings the predictions are key, so this presentation will focus on the developed of aggregated approaches that estimate the total uncertainty in predictions.

# Challenging features of residuals in hydrology

(Cotter River, ACT)



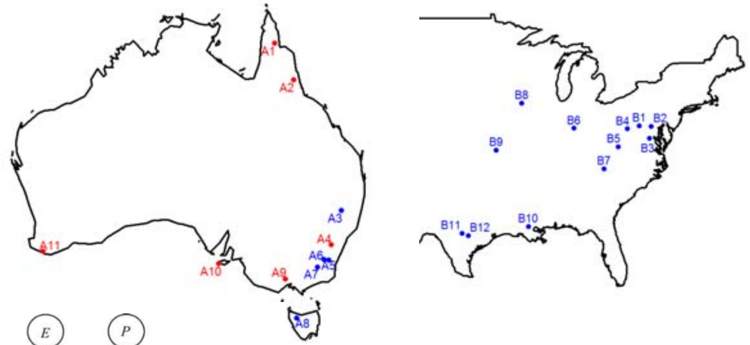

**Streamflow time series**

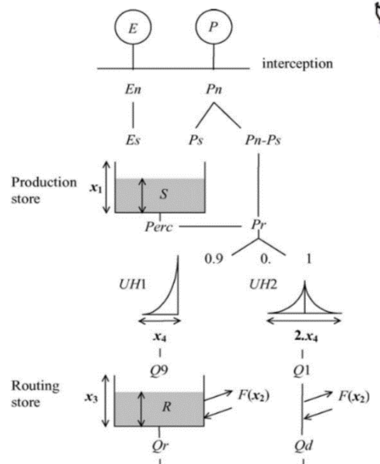
**Residual errors time series**

- **Errors are heteroscedastic** (larger errors in large flows)
- **Errors are persistence** (not independent between time steps)
- Appropriate representation of both "features" is required to achieve reliable probabilistic predictions

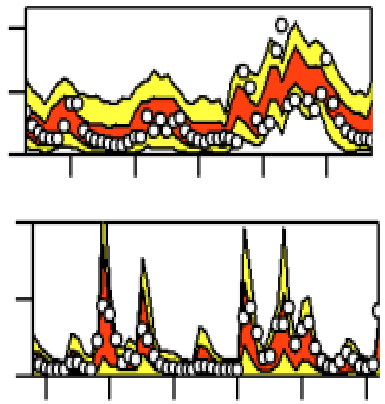
# What is the “best” residual error model for making daily streamflow probabilistic predictions?

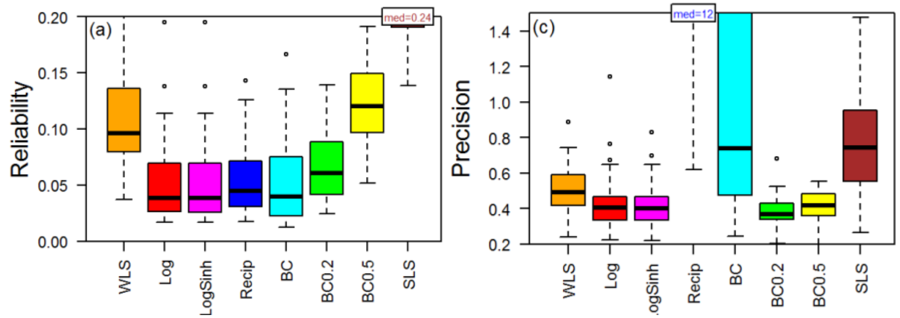
- Research Gap: No study had compared the wide range of residual error models
- 8 different residual error models
  - SLS, WLS, log, logsinh, Box-Cox transforms
- Comprehensive comparisons
  - 23 catchments from Australia and USA
  - 2 hydrological models (GR4J, HBV)
- Cross-validation over 10 yr period
  - 3500+ model calibrations
  - 4000+ CPU hours (150 days) on HPC
- Multiple performance metrics
  - Reliability, precision and bias
- McInerney et al (WRR2017)




[Perrin *et al,* 2003]



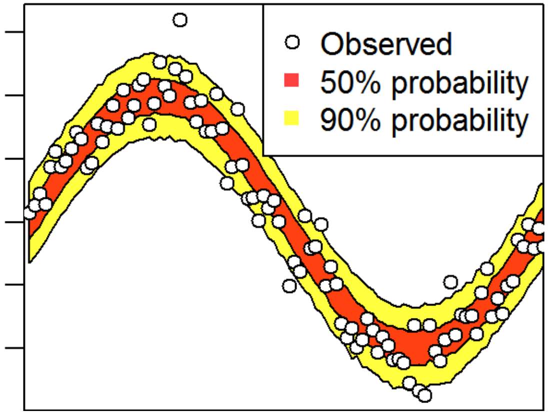


# What makes good probabilistic predictions?

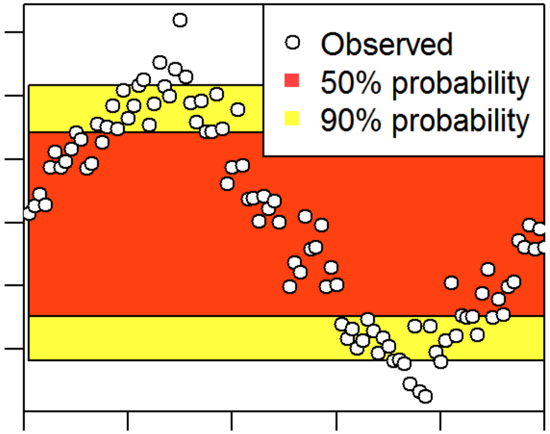
We want predictions that are

- **Reliable:** Predictions statistically consistent with observed data
- **Precise:** Small uncertainty in predictions
- With **low volumetric bias:** total volume from predicted flow matches observations

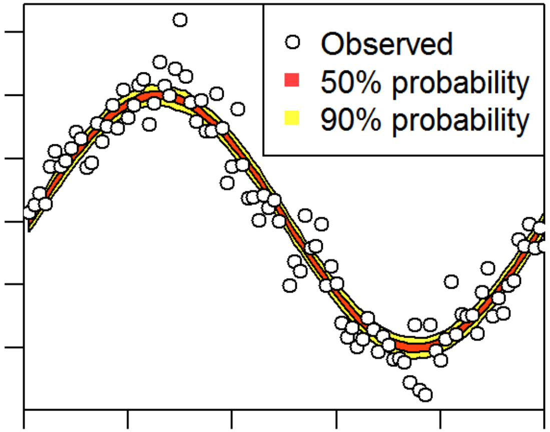
**Reliable, precise, unbiased**



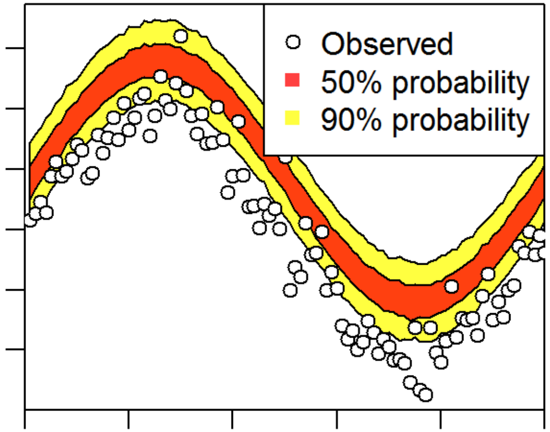

**Reliable but imprecise**




**Precise but unreliable**



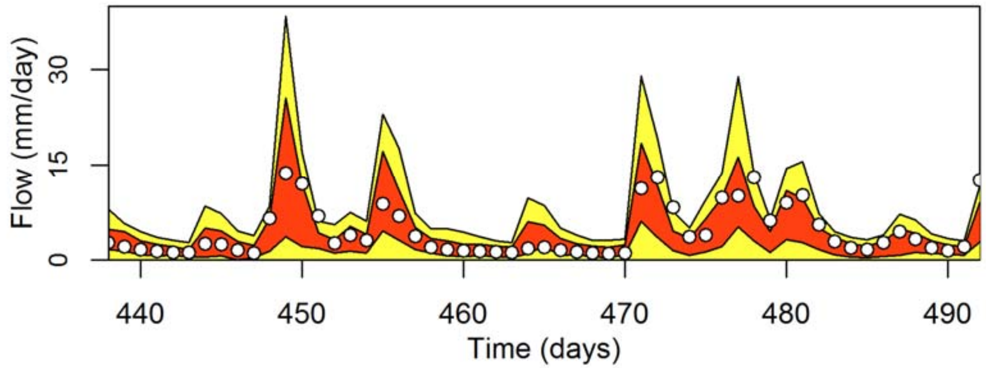

**Biased**

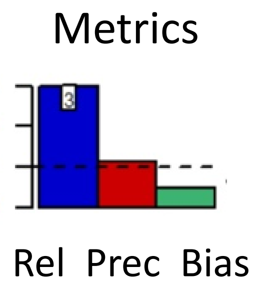


# Error model has large impact on probabilistic predictions: Perennial Catchments

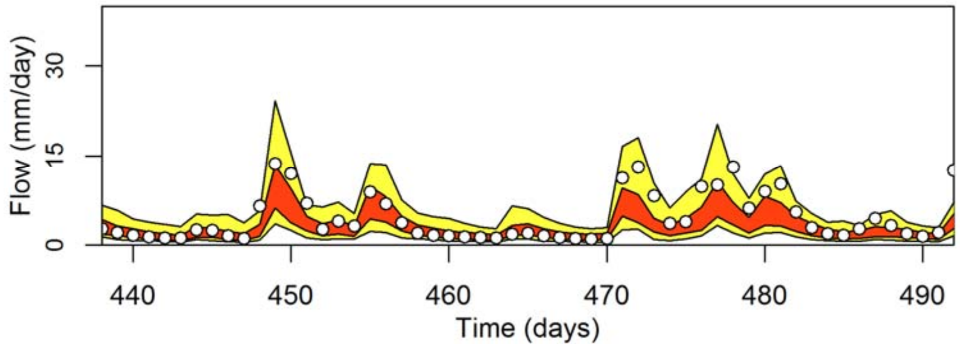
- Perennial catchment (Spring River, USA), GR4J hydro model

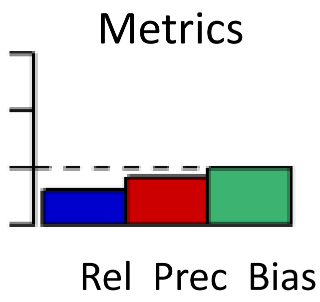
**Weighted Least Squares (WLS)**




**Log transformation**



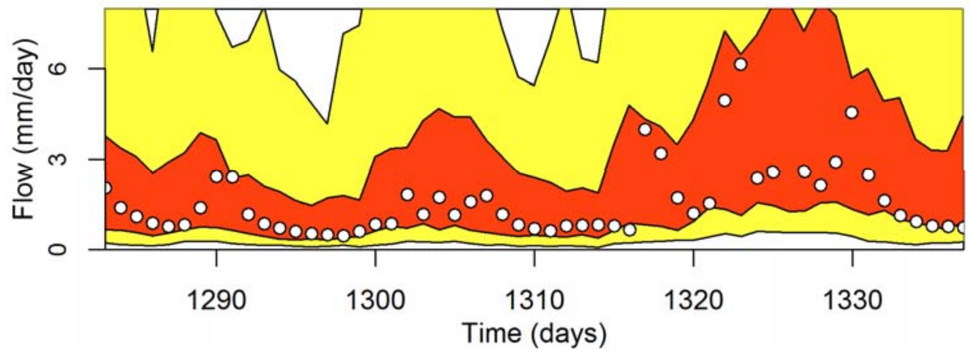



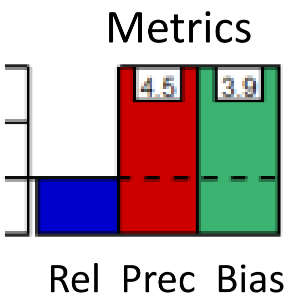

- Weight least squares has poor reliability and worse precision
- Log transformation better reliability and precision
- Transformational approaches (Log and BC) better handle skew in observed residuals

# Error model has large impact on probabilistic predictions: Ephemeral Catchments

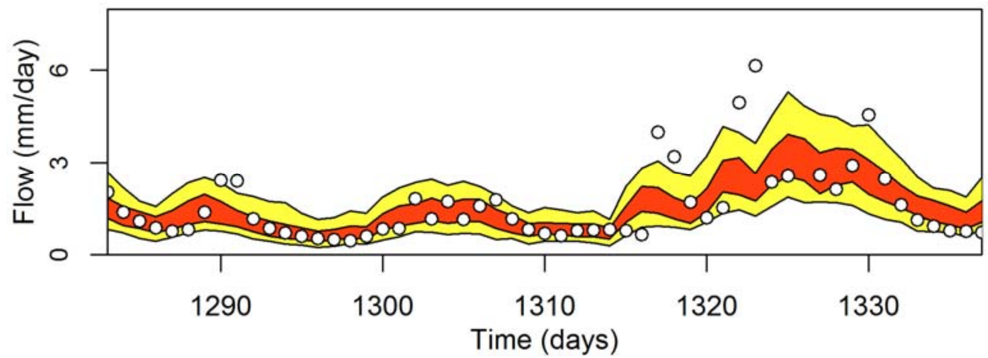
- Ephemeral catchment (Rocky River, SA), HBV hydro model

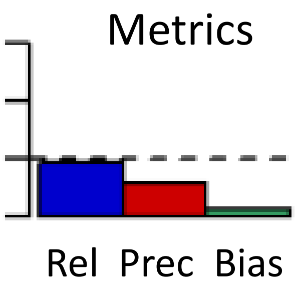
**Log transformed flows**



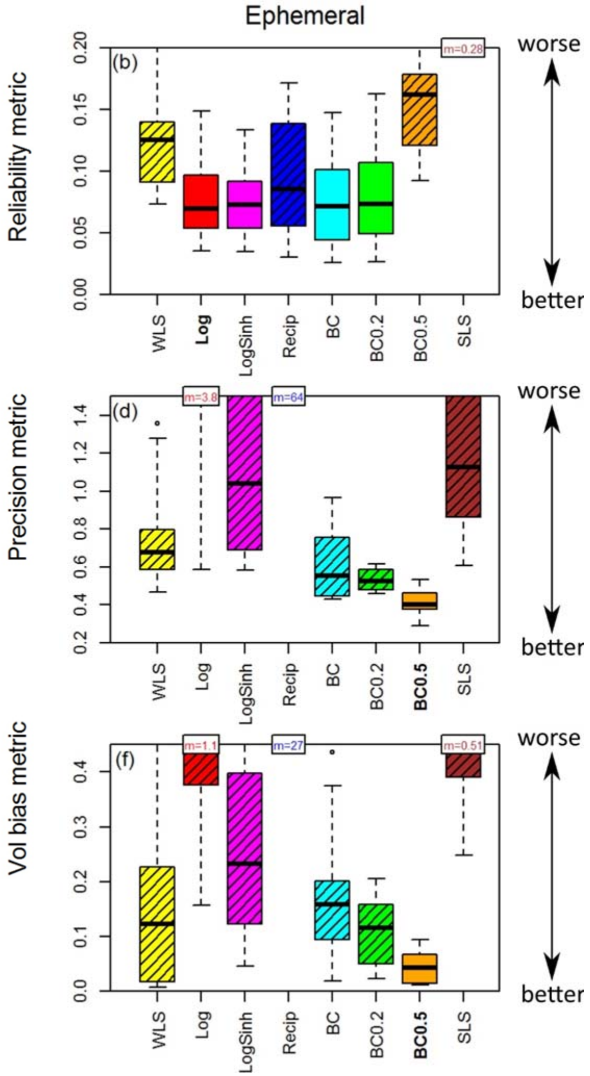

**Box Cox transformed flows**

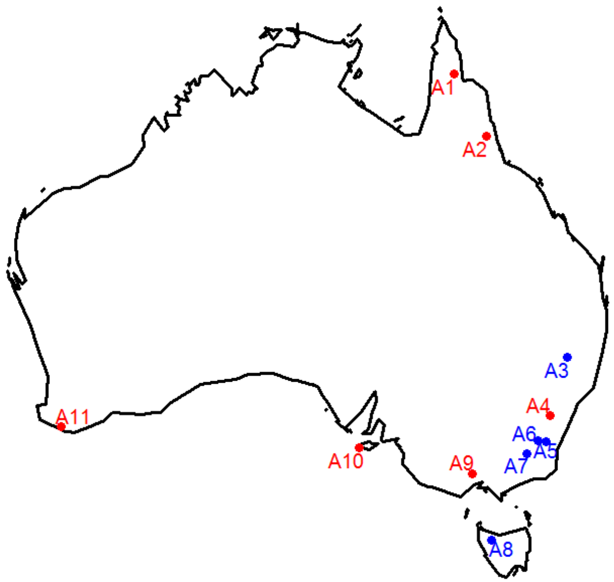



- Log produces poor precisions (unrealistically large uncertainty) and large bias in ephemeral catchments
- Box Cox transformation (lambda=0.2) performs much better
- BC transformation better handles zero flows in ephemeral catchments

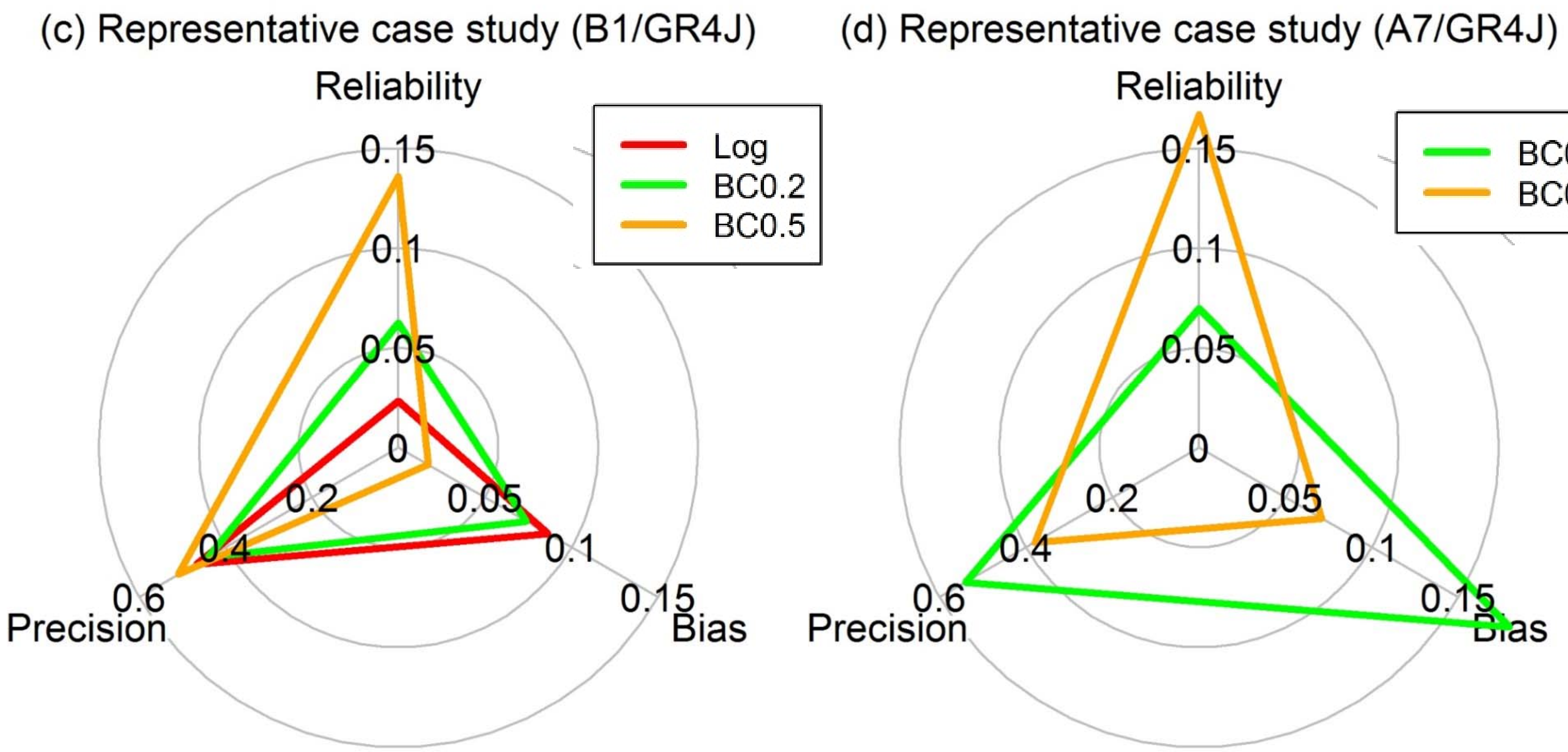
# Improvements in probabilistic performance largest in ephemeral catchments




- Improved reliability
- Improved precision 105% to 40% of obs streamflow
- Reduced bias from 25% to 4%

# Can't always get everything that we want




- In most catchments, tradeoffs between reliability, precision and bias seen even amongst the "best" error models
- Further investigation underway to reduce tradeoffs

# Key Insights

- Error models based on Box-Cox transform produce more reliable predictions
- Choosing the best error models depends on the type of flow regime in the catchment.
- More complex error models do not necessarily produce the best predictions
- No single error model performs best in all aspects of predictive performance

# Broad Recommendations

In **perennial** catchments, use

- Log error model if reliability is important
- Box Cox transformation with $\lambda=0.2$ if precision is important
- Box Cox transformation with $\lambda=0.5$ if low bias is important

In **ephemeral** catchments, use

- Box Cox transformation with $\lambda=0.2$ if reliability is important
- Box Cox transformation with $\lambda=0.5$ if precision/bias important

Does not include an in-depth analysis of performance trade-offs

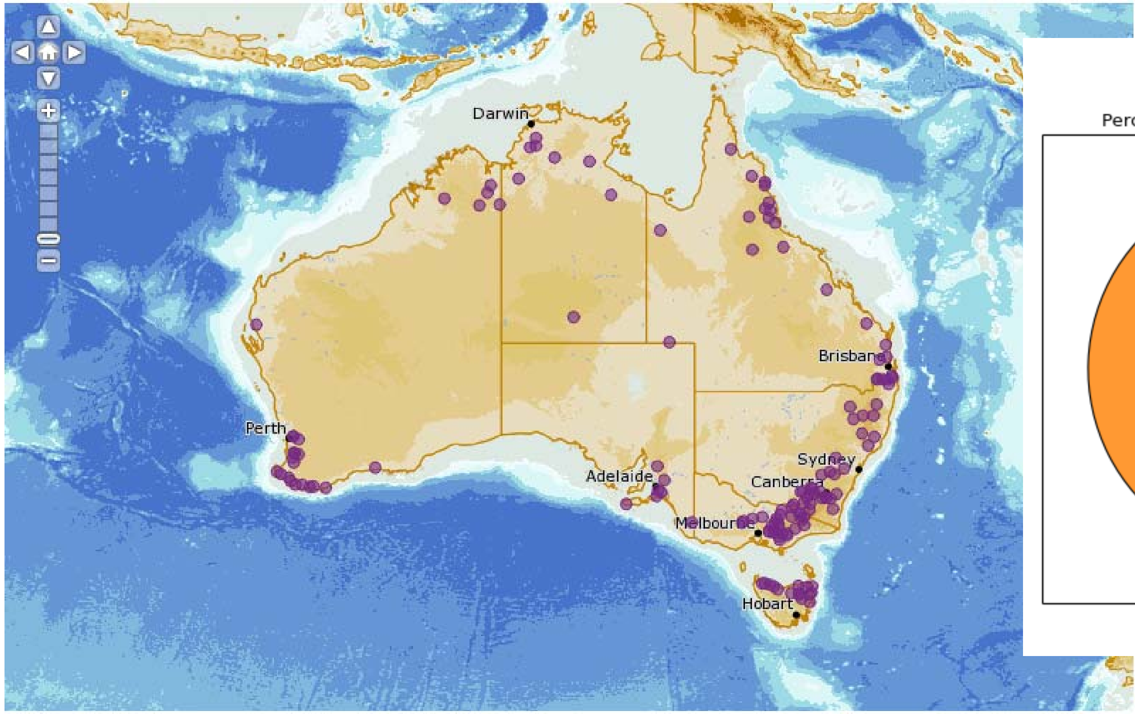
# Summary

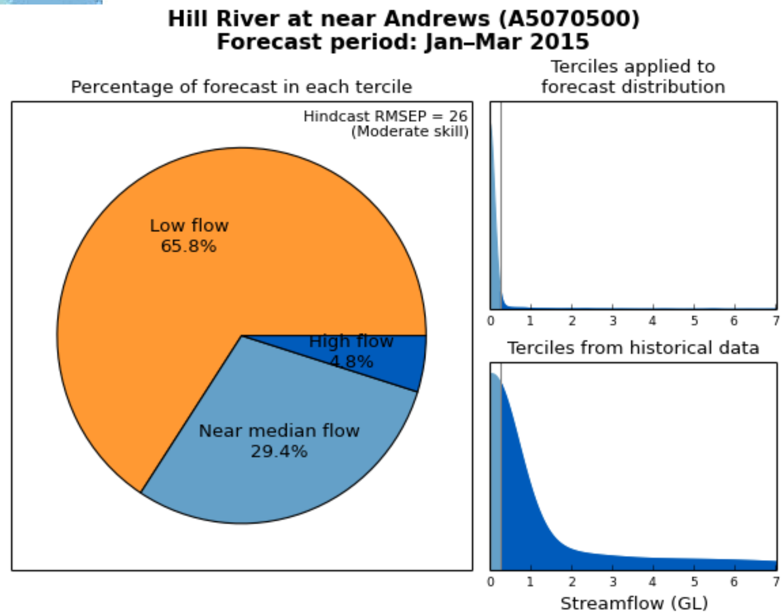
- Comprehensive evaluation a range of approaches for modelling total predictive uncertainty
  - Eight Approaches: Simple=>Complex
  - Empirical results: 23 catchments and 2 hydro models
  - Theory: Understanding when and why approaches provide good or bad predictive performance, e.g. ephemeral versus perennial catchments
- Broad recommendations
  - Selection of error models to use in different catchment types
  - Achieve reliable and precise probabilistic predictions
- Practical implications: Simplest is often best!
  - Prudent use of simple approaches => best predictive performance
  - Simple to implement for practitioners
  - Further work ongoing to reduce performance tradeoffs

# Outline

- Improving Hydrological Predictions
  - Overview of advances in flexible modelling, in lumped and spatial outcomes
  - Improving model calibration
  - Spatial Rainfall Modelling
- Improving streamflow probabilistic predictions
  - Practical guidance on which residual error model to improve reliability and precision of probabilistic predictions
- Impacts on Forecasting

# Bureau of Meteorology
# Seasonal Streamflow Forecasts

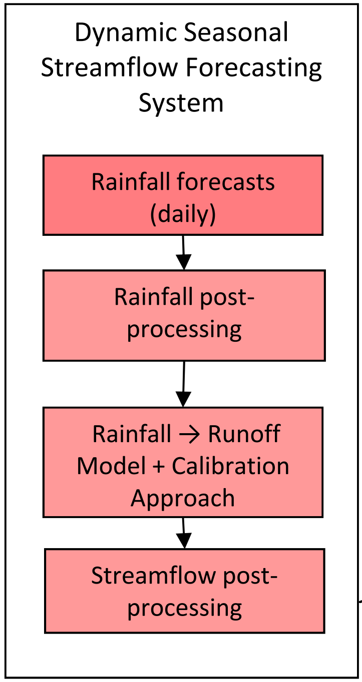



- Seasonal forecasts at ~300 locations
- Used upon by large number of water managers around Australia
- Hydrological forecasts have wide range of uncertainty
- BOM is using our techniques to characterize uncertainty and ultimately improve probabilistic predictions

# Impacts on Forecasting: Streamflow post processing at monthly time step

- Strategic Objectives of Bureau’s Seasonal Streamflow Forecasting (SSF) System
  - Aims: Provide forecasts across a wide range of Australian sites




Enhanced streamflow post-processor at monthly time scale

Outcomes:

- Majority of sites with reliable forecasts
- Majority of sites with forecast “sharpness” better than climatology
- Operationalised within Bureau’s system

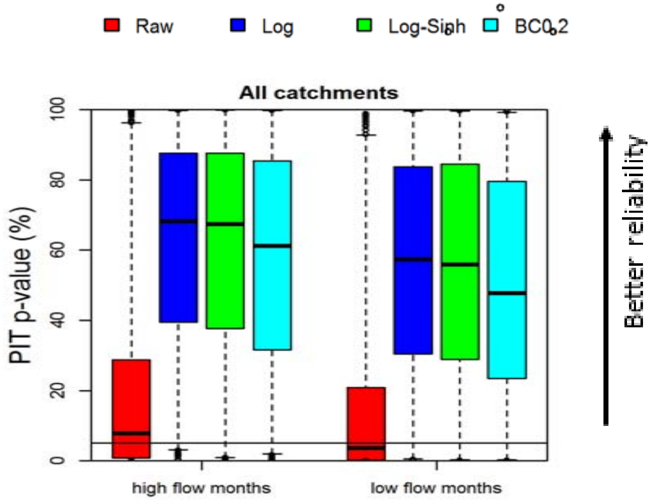
# Enhance Streamflow Postprocessor at Monthly and Seasonal Time scale

- Evaluated different residual models for post-process streamflow forecasts at monthly time scale
  - Raw: No Post-Processing
  - Log, logsinh and BC0.2
- Evaluation on wide range of 300+ catchment across Australia
  - Classified into dry and wet using Aridity Index <0

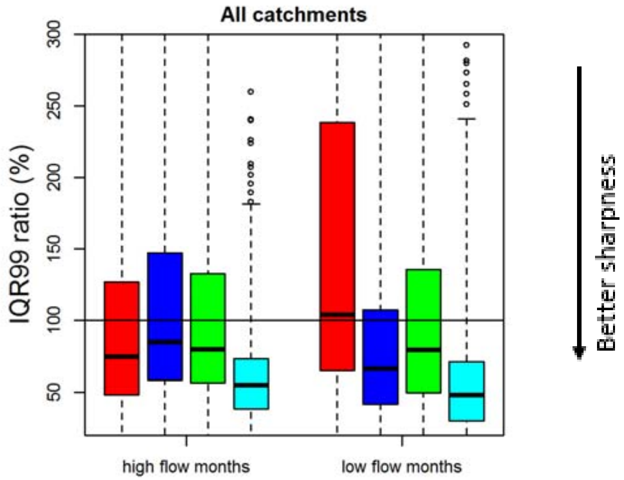
- Multiple metrics
  - Reliability (PIT plot p-value)
  - Sharpness (Inter quantile range for 99$^{th}$ percentile)
    - Expressed as % of climatology IQR
    - > 100% sharpness is worse than climatology
    - < 100% sharpness is better than climatology
  - CRPS SS

Woldemsekel F., Lerat, J., Tuteja, N., Shin, D.H., Thyer, M., McInerney, D., Kavetski, D., Kuczera, G. (2017) Evaluating residual error approaches to post-processing monthly and seasonal streamflow forecasts, *Hydrology and Earth System Sciences* (in preparation).

# Outcomes for Monthly Forecasts: Better reliability and sharpness



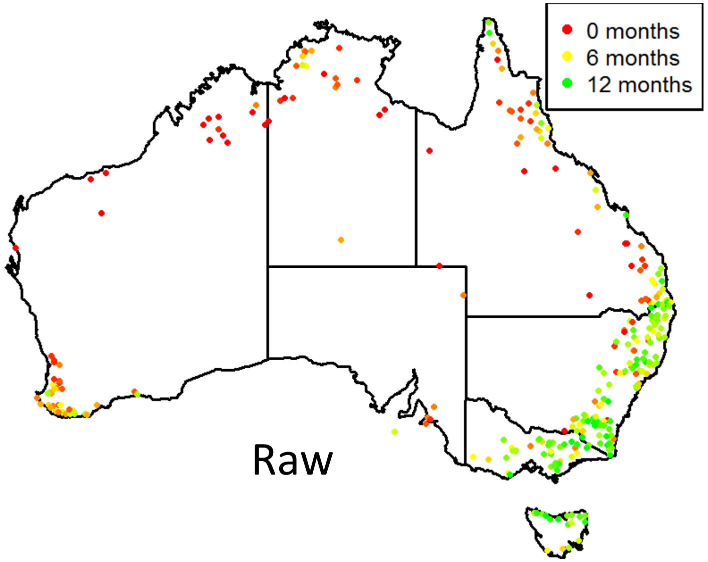

- Post-processing provides more sites with reliable forecasts: Increase from ~50% to >90% sites
- Similar results for CRPS



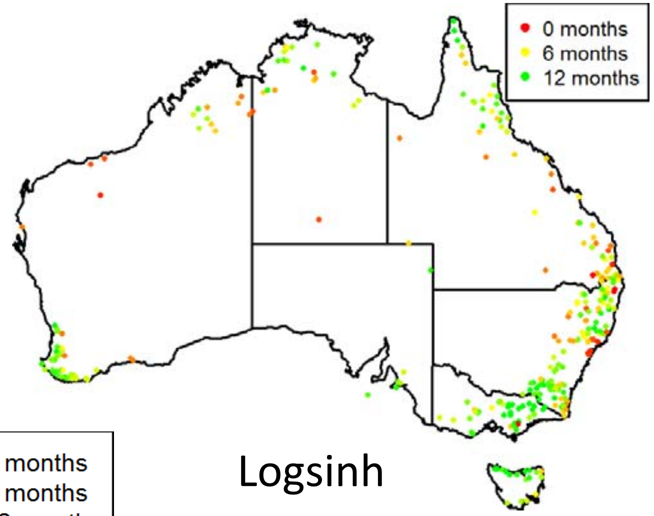

- BC0.2 provides more sites with sharpness better than climatology: Increase from 60% of sites to ~90% sites

Woldemsekel F., Lerat, J., Tuteja, N., Shin, D.H., Thyer, M., McInerney, D., Kavetski, D., Kuczera, G. (2017) Evaluating residual error approaches to post-processing monthly and seasonal streamflow forecasts, *Hydrology and Earth System Sciences* (in preparation).

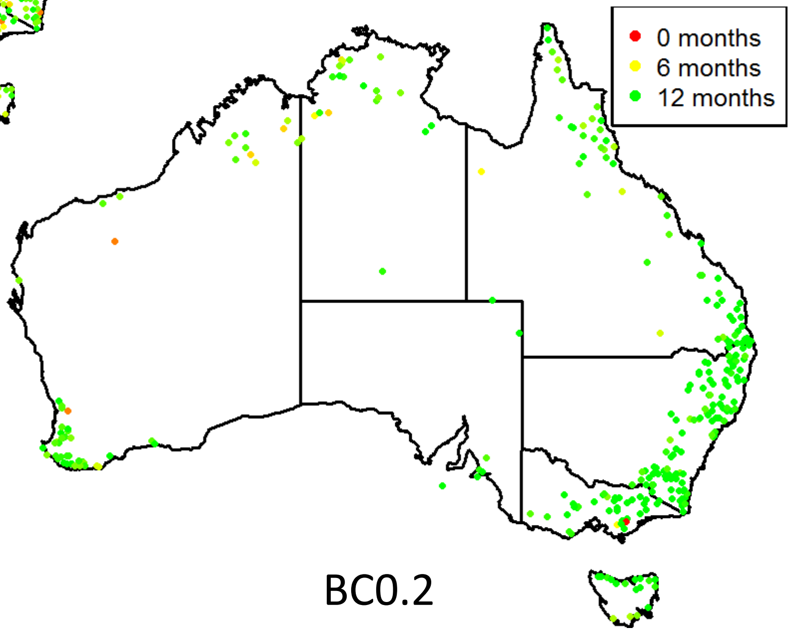
# Outcomes for Monthly Forecasts: Better reliability and sharpness




Raw



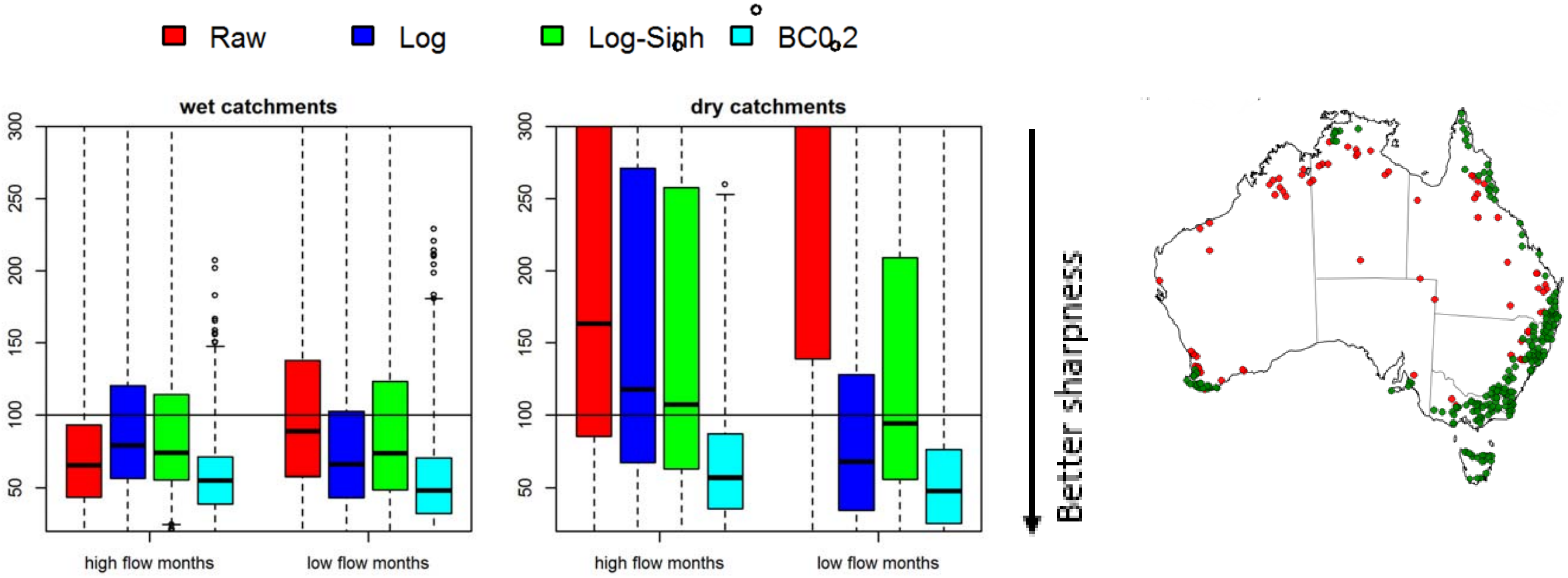

Logsinh




BC0.2

- No. of months at each site with forecasts that are reliable and with "sharpness" better than climatology.

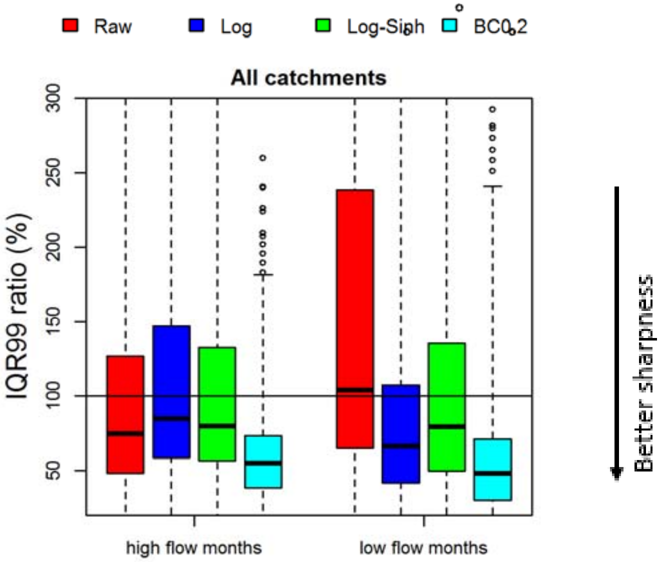
# Major increase in sharpness in dry catchments



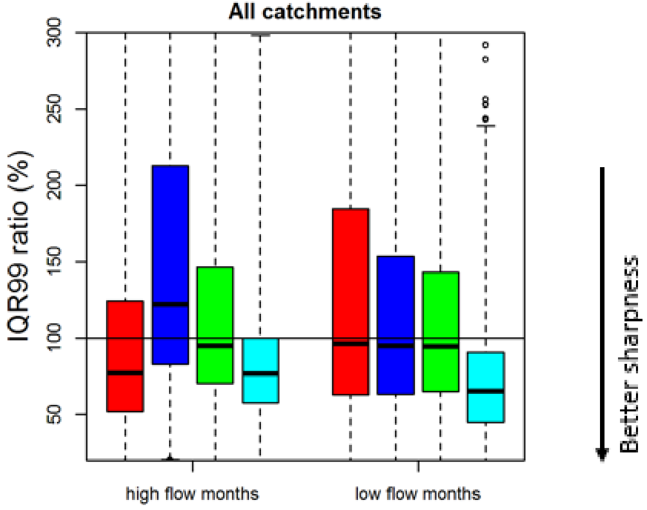

- In dry catchments, high flow months
  - log and logsinh models have >50% of sites worse then climatology
  - BC0.2 only model that provides more (80%) sites with sharpness better than climatology
  - Little change in reliability or CRPS

Woldemsekel F., Lerat, J., Tuteja, N., Shin, D.H., Thyer, M., McInerney, D., Kavetski, D., Kuczera, G. (2017) Evaluating residual error approaches to post-processing monthly and seasonal streamflow forecasts, *Hydrology and Earth System Sciences* (in preparation).

# Improvements in sharpness are bigger for seasonal forecast than monthly forecasts




Monthly forecasts




Seasonal forecasts

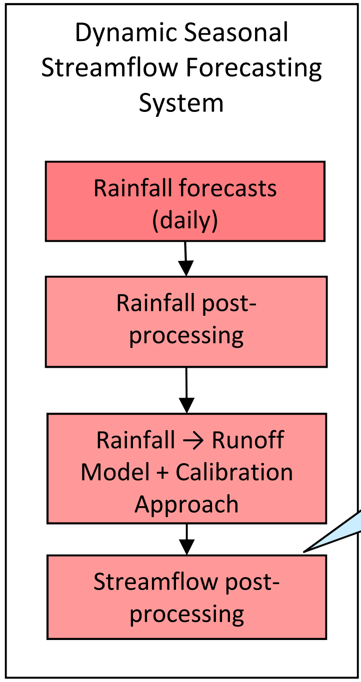
- Seasonal forecast,
  - Log and logsinh: 50% or less sites better climatology
  - BC0.2: >80% sites better climatology

- No difference between models for CRPS and Reliability

Woldemsekel F., Lerat, J., Tuteja, N., Shin, D.H., Thyer, M., McInerney, D., Kavetski, D., Kuczera, G. (2017) Evaluating residual error approaches to post-processing monthly and seasonal streamflow forecasts, *Hydrology and Earth System Sciences* (in preparation).

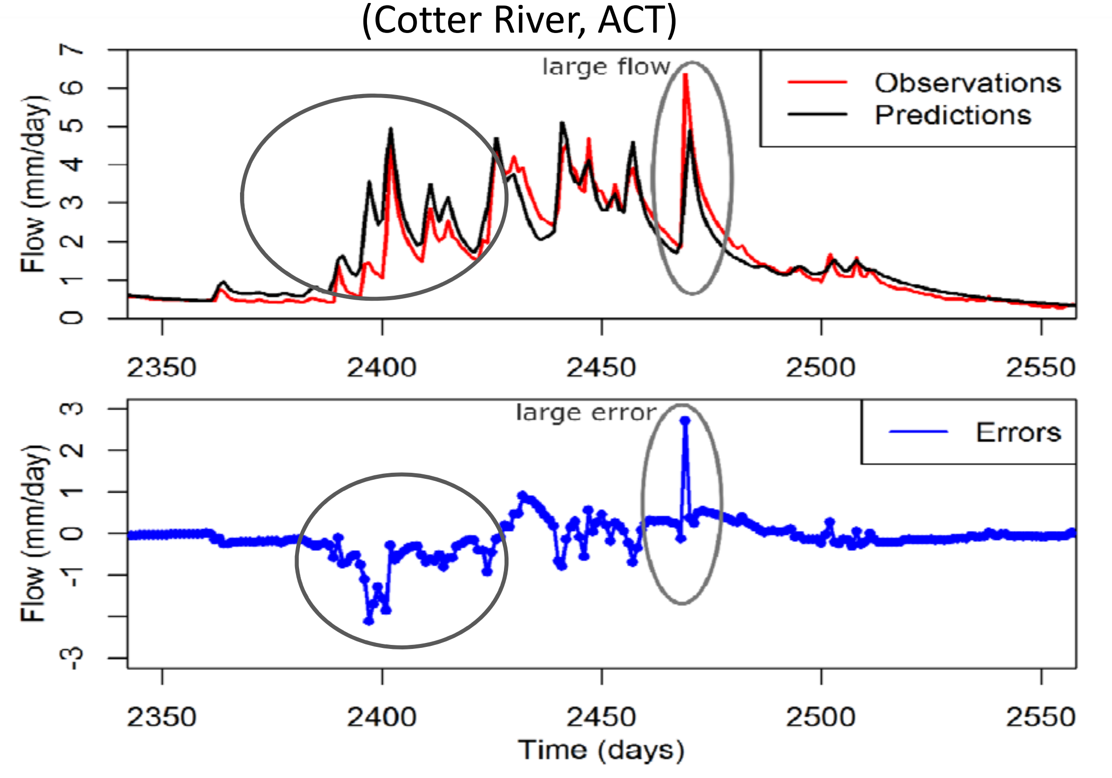
# Impact on Forecasts: Towards Seamless Forecasts

- Strategic Objectives of Bureaus Seasonal Streamflow Forecasting (SSF) System
  - Current aims: Provide “seamless” forecasts (seamless = single product able to be aggregated in time from daily to monthly scale)
- Progress towards this objective




Enhanced daily streamflow post-processing (obs. rain)
**Outcomes:** Incorporating persistence improves reliability when aggregating from daily to monthly time scale

# Challenging features of residuals in hydrology



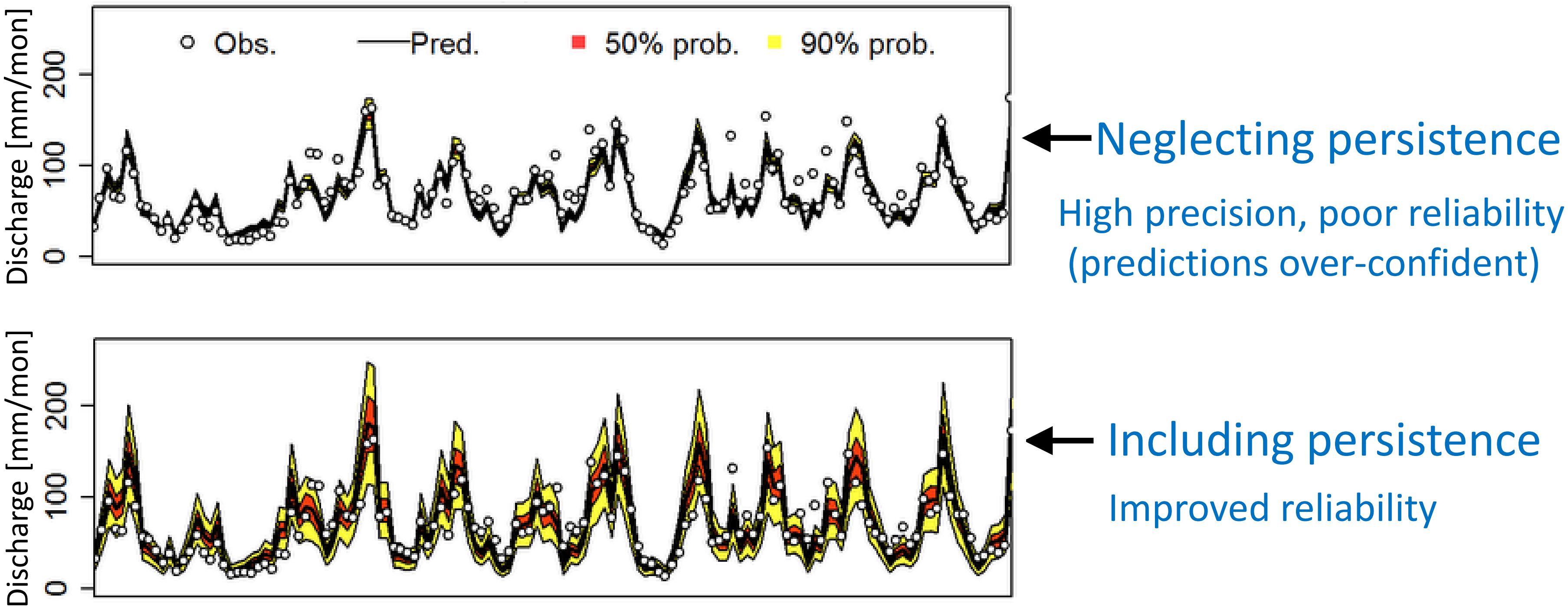

**Streamflow time series**

**Residual errors time series**

- **Errors are heteroscedastic** (larger errors in large flows)
- **Errors are persistence** (not independent between time steps)
- Appropriate representation of both "features" is required to achieve reliable probabilistic predictions

# Importance of modelling persistence in errors

- Persistence important when aggregating data
  - E.g. daily predictions aggregated to monthly values

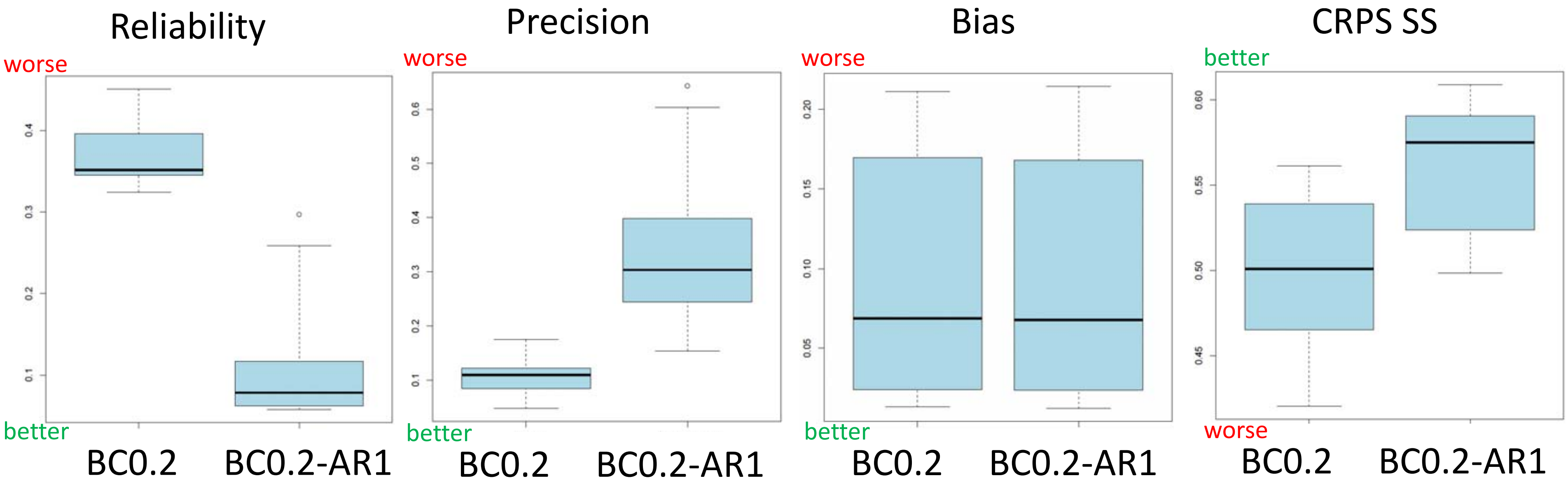



[Evin et al, 2014]

- Ignoring persistence produces under-estimation of predictive uncertainty when aggregating data => under-estimates risks of extreme events

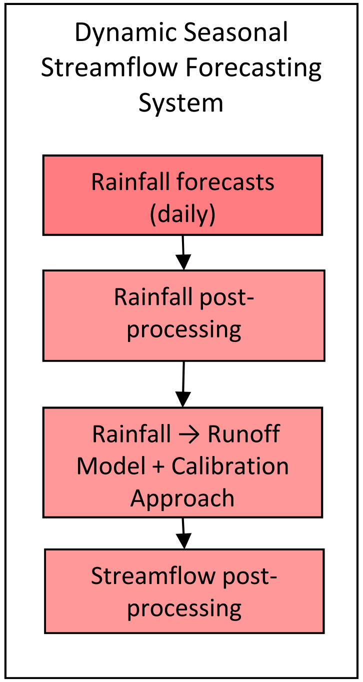
# Modelling persistence critical when aggregating daily data


## Monthly metrics




- Without Persistence (BC0.2), grossly over-confident: good precision but very unreliable => low CRPS SS
- With Persistence (BC0.2+AR1), credible predictions: lower precision but better reliability => high CRPS SS

# Summary: Impacts on Forecasting




1. Enhanced streamflow post-processor at monthly time scale

- **Outcomes:**
  - Majority of sites with reliable forecasts
  - Majority of sites with forecast "sharpness" better than climatology
  - Operationalised within Bureau's system
- Contribution towards Bureaus objective of reliable forecasts at wide range of sites across Australia

2. Enhanced daily streamflow post-processing (obs. rain)

- **Outcomes:**
  - Incorporating persistence improves reliability when aggregating from daily to monthly time scale
  - Currently being operationalised
- Contribution towards Bureaus objective of seamless forecasts across a range of time scales (daily to monthly)

# Questions?

## Further Information

Email: mark.thyer@adelaide.edu.au or connect with me on Linkedin

Subscribe to Intelligent Water Decisions Research Group Blog

www.waterdecisions.org/blog/

## References

Bennett, B., M. Thyer, M. Leonard, M. Lambert, and B. Bates (2016), A comprehensive and systematic evaluation framework for a parsimonious daily rainfall field model, *J Hydrol*10.1016/j.jhydrol.2016.12.043

Evin, G., D. Kavetski, M. Thyer, and G. Kuczera (2013), Pitfalls and improvements in the joint inference of heteroscedasticity and autocorrelation in hydrological model calibration, *Water Resour Res*, *49*(7), 4518-4524, http://dx.doi.org/10.1002/wrcr.20284

Evin, G., M. Thyer, D. Kavetski, D. McInerney, and G. Kuczera (2014), Comparison of joint versus postprocessor approaches for hydrological uncertainty estimation accounting for error autocorrelation and heteroscedasticity, *Water Resour Res*, *50*(3), 2350-2375, http://dx.doi.org/10.1002/2013WR014185

Fenicia, F., D. Kavetski, H. H. G. Savenije, and L. Pfister (2016), From spatially variable streamflow to distributed hydrological models: Analysis of key modeling decisions, *Water Resour Res*, *52*(2), 954-989, 10.1002/2015WR017398

Kavetski, D., and F. Fenicia (2011), Elements of a flexible approach for conceptual hydrological modeling: 2. Application and experimental insights, *Water Resour. Res.*, *47*(11), W11511, 10.1029/2011wr010748

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

Woldemsekel F., Lerat, J., Tuteja, N., Shin, D.H., Thyer, M., McInerney, D., Kavetski, D., Kuczera, G. (2017) Evaluating residual error approaches to post-processing monthly and seasonal streamflow forecasts, *Hydrology and Earth System Sciences* (in preparation).