Feature-specific verification of ensemble forecasts

www.cawcr.gov.au

Beth Ebert
CAWCR Weather & Environmental Prediction Group

Australian Government
Bureau of Meteorology

The Centre for Australian Weather and Climate Research
A partnership between CSIRO and the Bureau of Meteorology

CSIRO

Uncertainty information in forecasting

Operational Forecaster Uncertainty Needs and Future Roles

DAVID R. NOVAK

NOAA/National Weather Service, Eastern Region Headquarters, Scientific Services Division, Bohemia, New York

DAVID R. BRIGHT

NOAA/National Weather Service, Storm Prediction Center, Norman, Oklahoma

MICHAEL J. BRENNAN

NOAA/National Weather Service, Hydrometeorological Prediction Center, Camp Springs, Maryland

(Manuscript received 20 March 2008, in final form 6 June 2008)

ABSTRACT

Key results of a comprehensive survey of U.S. National Weather Service operational forecast managers concerning the assessment and communication of forecast uncertainty are presented and discussed. The survey results revealed that forecasters are using uncertainty guidance to assess uncertainty, but that limited data access and ensemble underdispersion and biases are barriers to more effective use. Some respondents expressed skepticism as to the added value of formal ensemble guidance relative to simpler approaches of estimating uncertainty, and related the desire for feature-specific ensemble verification to address this skepticism. Respondents reported receiving requests for uncertainty information primarily from sophisticated users such as emergency managers, and most often during high-impact events. The largest request for additional training material called for simulator-based case studies that demonstrate how uncertainty information should be interpreted and communicated.

Respondents were in consensus that forecasters should be significantly involved in the communication of uncertainty forecasts; however, there was disagreement regarding if and how forecasters should adjust objective ensemble guidance. It is contended that whether forecasters directly modify objective ensemble guidance will ultimately depend on how the weather enterprise views ensemble output (as the final forecast or as a guidance supporting conceptual understanding), the enterprise's commitment to provide the nec-

December 2008 *WAF*

For high impact events, forecasters want to see evidence that EPSs provide information that is as useful as deterministic forecasts and poor man's ensembles.

Forecasters indicated a desire for **feature-specific verification approaches** (e.g., Ebert and McBride 2000, Davis et al. 2006) to be applied to ensemble forecasts.

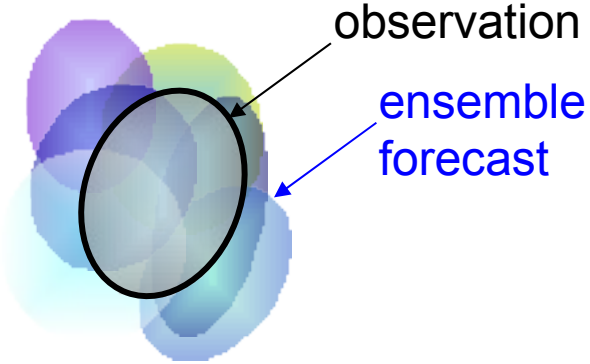
Verifying features in ensembles

Significant weather events can often be viewed as 2D "objects"

- tropical cyclones, heavy rain events, deep low pressure centres
- objects can be defined by an intensity threshold

What might the ensemble forecast look like?

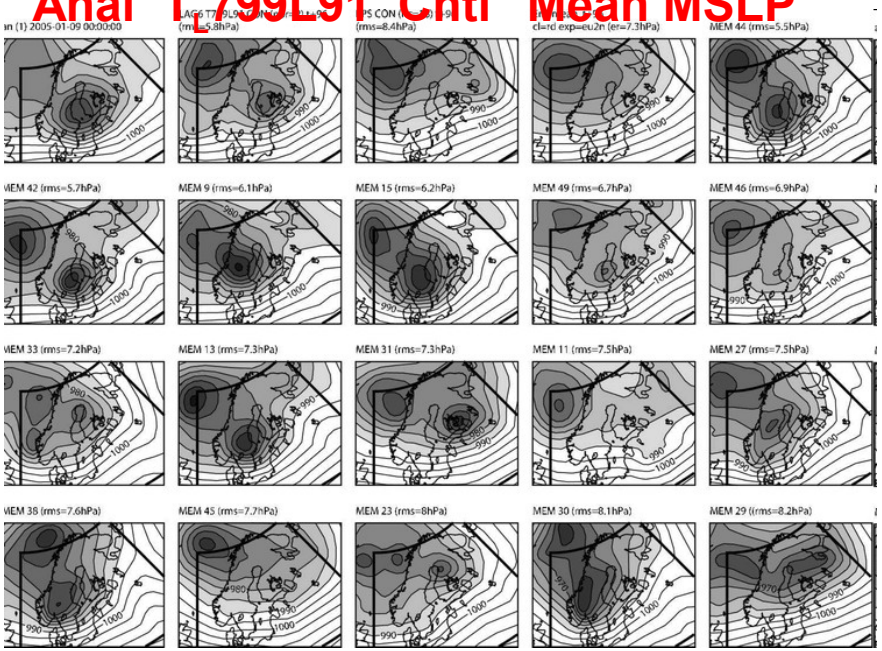
- spatial probability contour maps
- ensemble mean
- distributions of object properties (attributes)
  - location, size, intensity, etc.

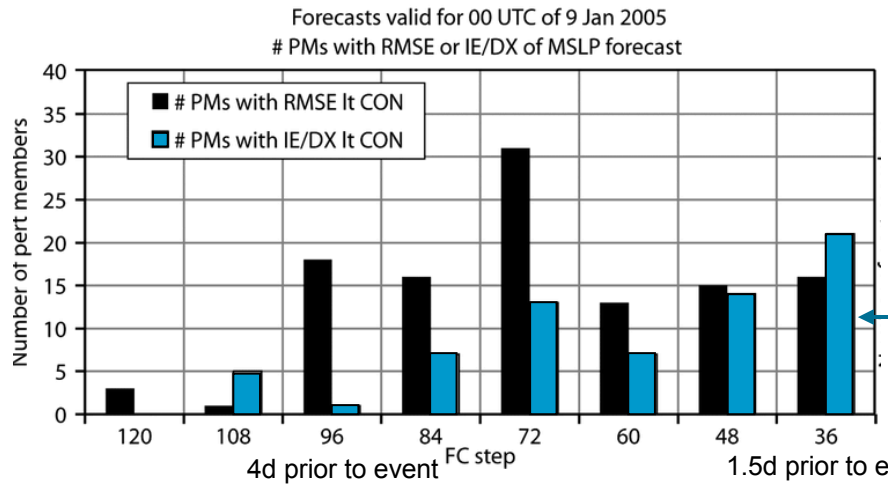


Strategies for verifying ensemble predictions of objects

1. Verify objects in probability maps
2. Verify "ensemble mean"
   - spatially averaged forecast objects (with histogram recalibration?)
   - generated from average object properties
3. Verify distributions of object properties
   - individual cases
   - many samples – use probabilistic measures

Buizza (MWR 2008)





Extreme event:

9 January 2005 storm (Gudrun)

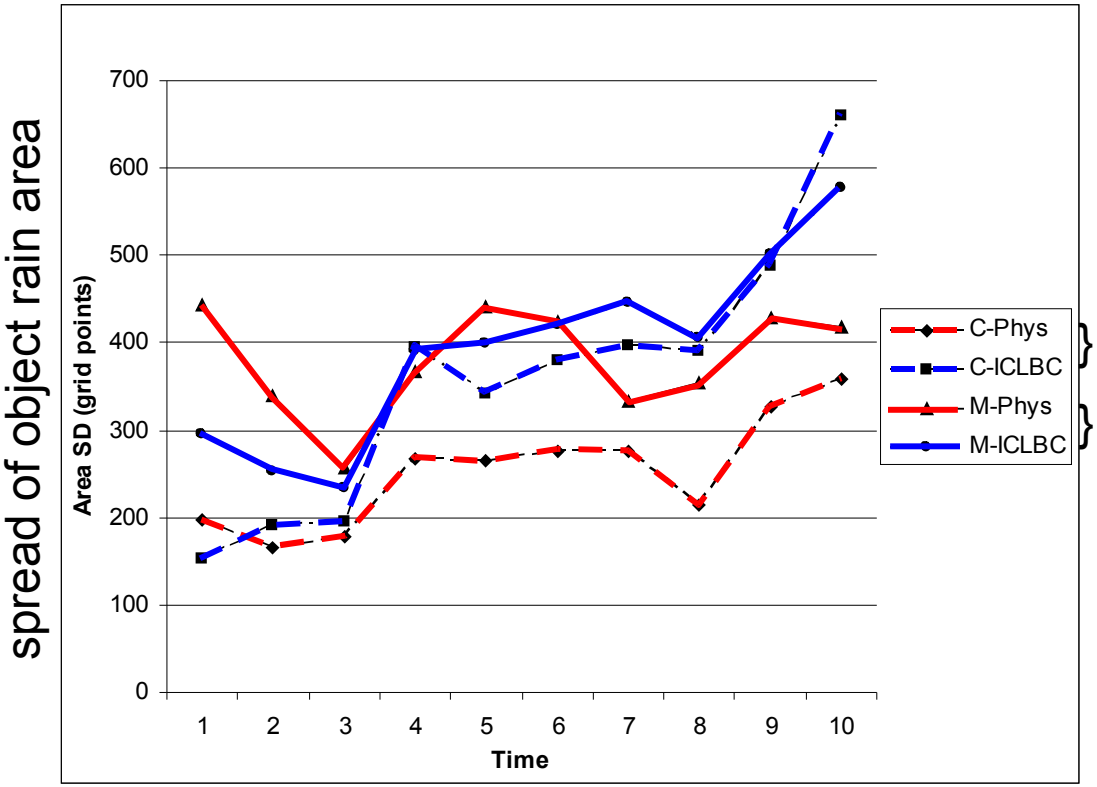
96h forecast from T_L399L62 EPS

Several members outperformed the control forecast

Better intensity and location

Gallus (2009) study

Compares forecasts of 6h rain from two 8-member ensembles ("Phys" and "ICLBC") using MODE and CRA object verification methods



Object verification shows:

- spread increases with time
- diurnal cycle

} CRA verification

} MODE verification

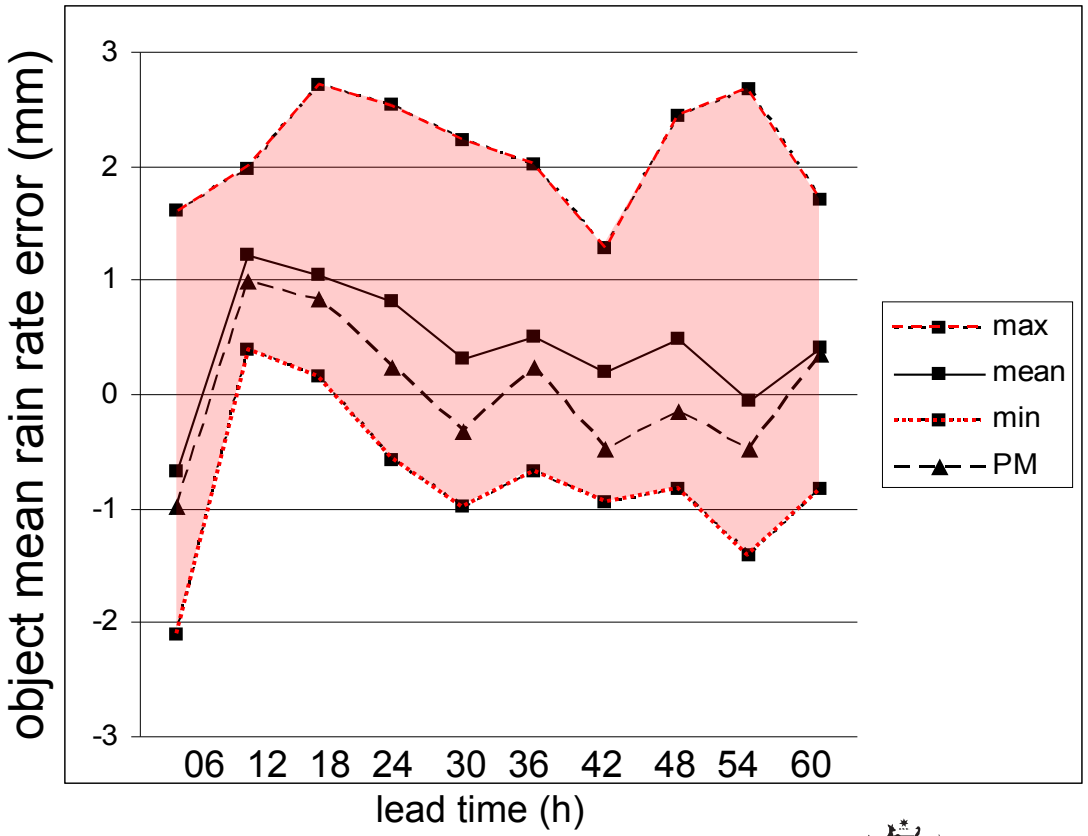
Australian Government
Bureau of Meteorology

The Centre for Australian Weather and Climate Research
A partnership between CSIRO and the Bureau of Meteorology

CSIRO

Gallus (2009) study

"Phys" ensemble verified using CRA method



Observed mean rain rate is enveloped within the ensemble at most lead times

Australian Government
Bureau of Meteorology

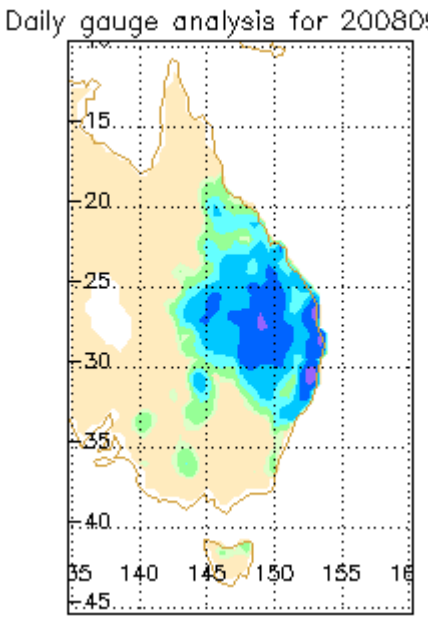
The Centre for Australian Weather and Climate Research
A partnership between CSIRO and the Bureau of Meteorology

CSIRO

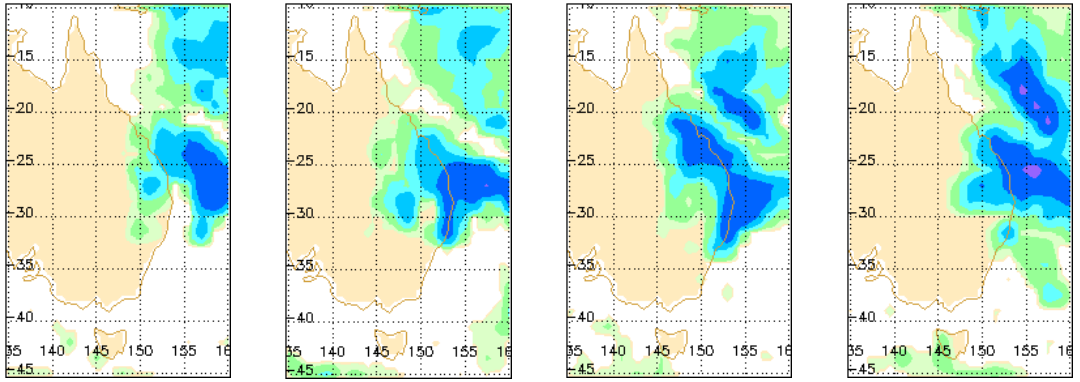
Practical questions to address

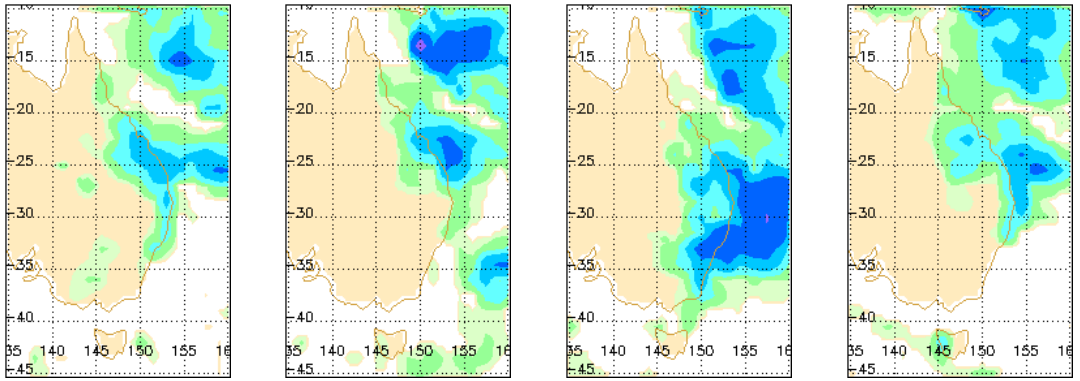
- What does the number of members predicting an event (feature) say about its likelihood?
- Is the spread of attributes (location, mean, max, area, volume) in the ensemble of events related to the uncertainty of the forecast event?
- Was the location of the center of the observed system found within an envelope (convex hull) of ensemble members' centers?
- How far was the observed feature center from the median location of the ensemble members?
- Was the observed heavy rain area within the range of the ensemble of areas? Volume? Maximum value?
- …

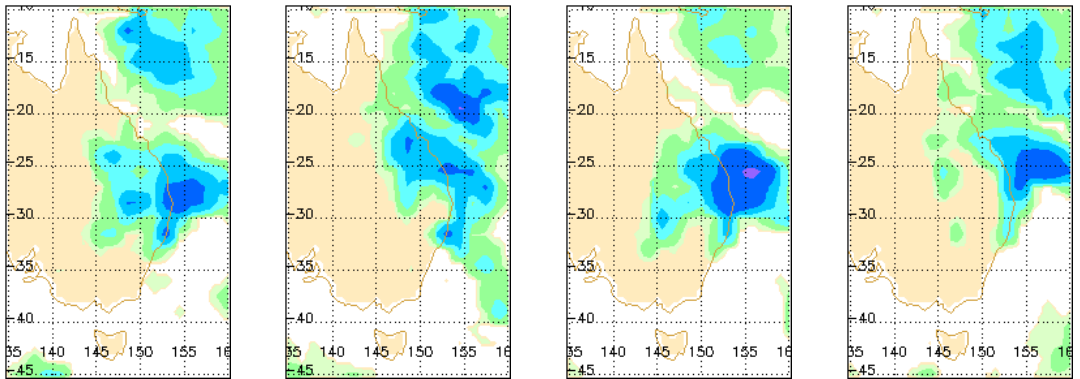
Example: East coast low, September 2008



Observed
00 UTC 5 Sept 2008





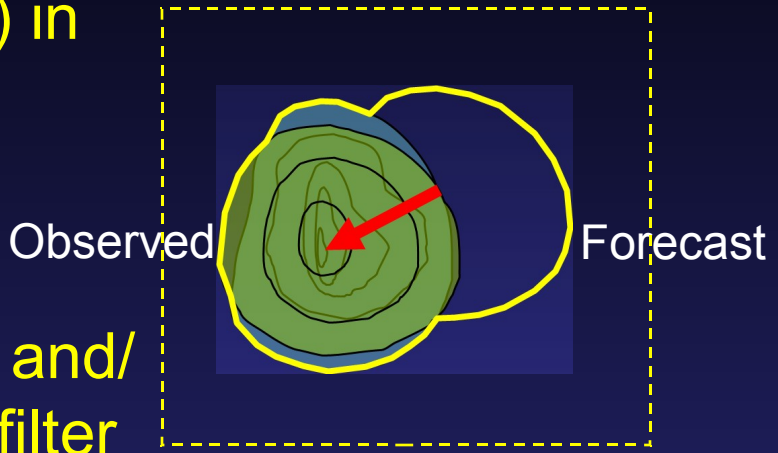


ECMWF EPS
4.5 day forecast
(first 12 of 51 members)

Contiguous Rain Area (CRA) verification

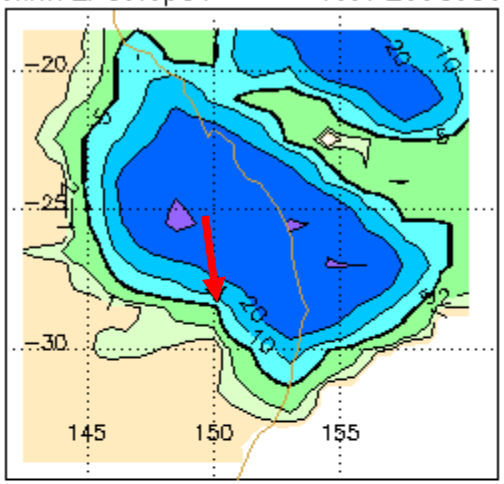
Ebert and McBride (*J. Hydrology,* 2000)

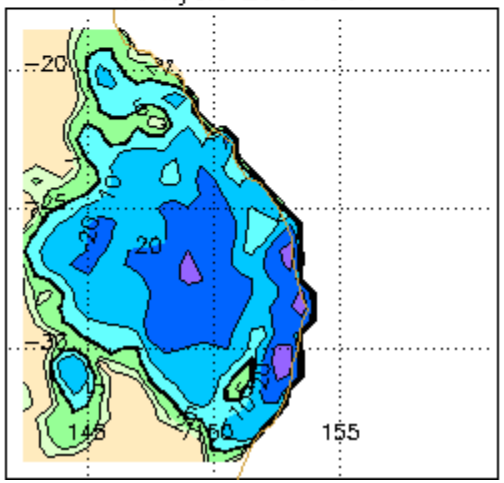
- Find Contiguous Rain Areas (CRA) in the fields to be verified
  - Take union of forecast and observations
  - Use minimum number of points and/ or total volume of parameter to filter out insignificant CRAs

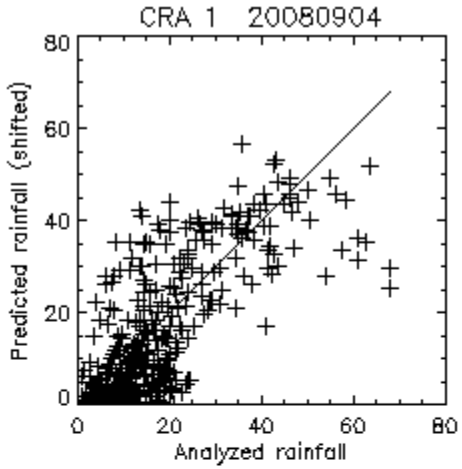


- Define a rectangular search box around CRA to look for best match between forecast and observations
- Displacement determined by shifting forecast within the box until MSE is minimized or correlation coefficient is maximized

CRA verification for an ensemble member







ECMWFEPS0.5p31 84-** fcst 20080904 n=387
(-33.50°,143.00°) to (-19.00°,160.00°)
Verif. grid=0.500° CRA threshold=5.0 mm/d

	Analysed	Forecast
# gridpoints ≥5 mm/d	358	233
Average rainrate (mm/d)	17.88	15.46
Maximum rain (mm/d)	68.07	56.66
Rain volume (km³)	19.12	16.53

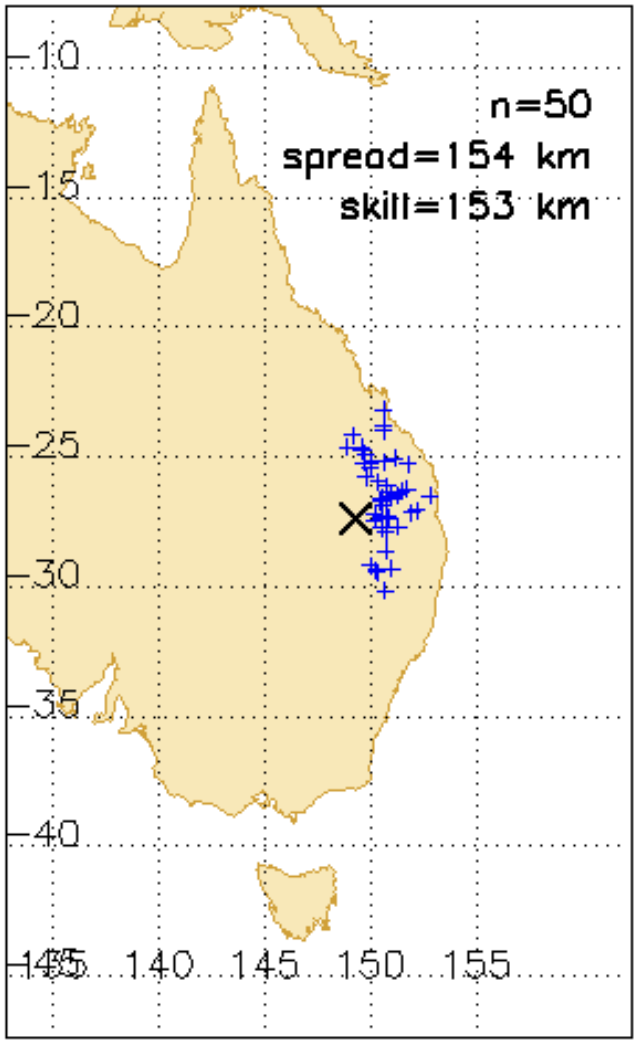
Displacement (E,N) = [-0.50°,3.00°]

	Original	Shifted
RMS error (mm/d)	17.84	10.29
Correlation coefficient	0.243	0.750

Error Decomposition:

Displacement error	66.7%
Volume error	1.8%
Pattern error	31.4%

Distribution of feature locations



ECMWF EPS 4.5 day forecast
valid 00 UTC 5 Sept 08

location = mass weighted center of event

spread = average distance to ensemble median location

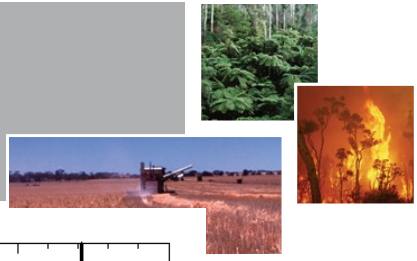
skill = distance between ensemble median and observed location

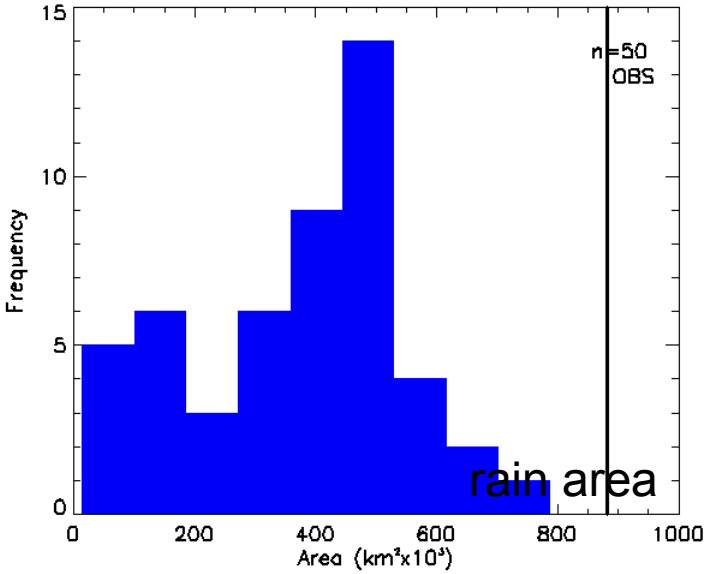
Australian Government
Bureau of Meteorology

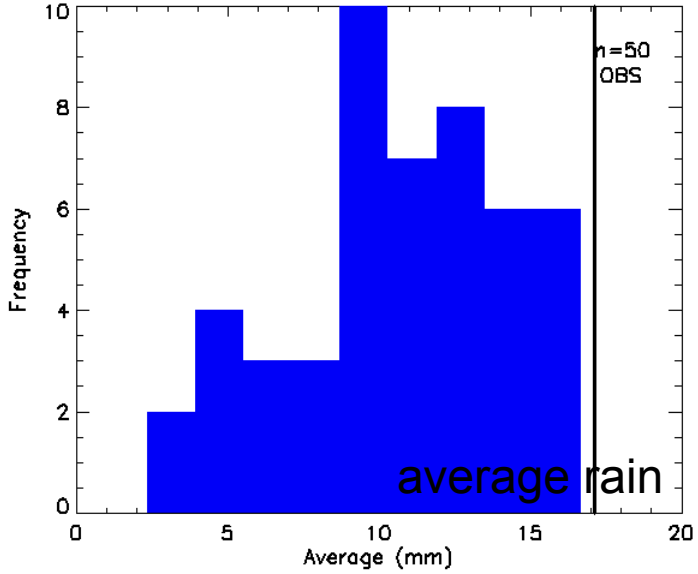
The Centre for Australian Weather and Climate Research
A partnership between CSIRO and the Bureau of Meteorology

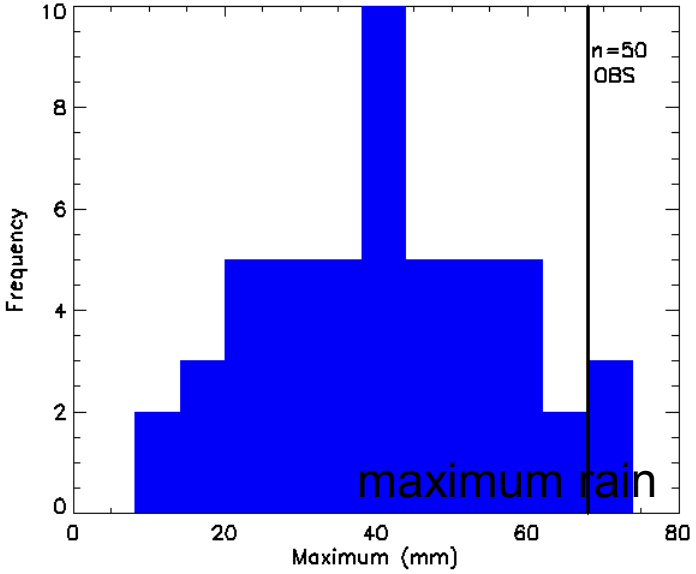
CSIRO

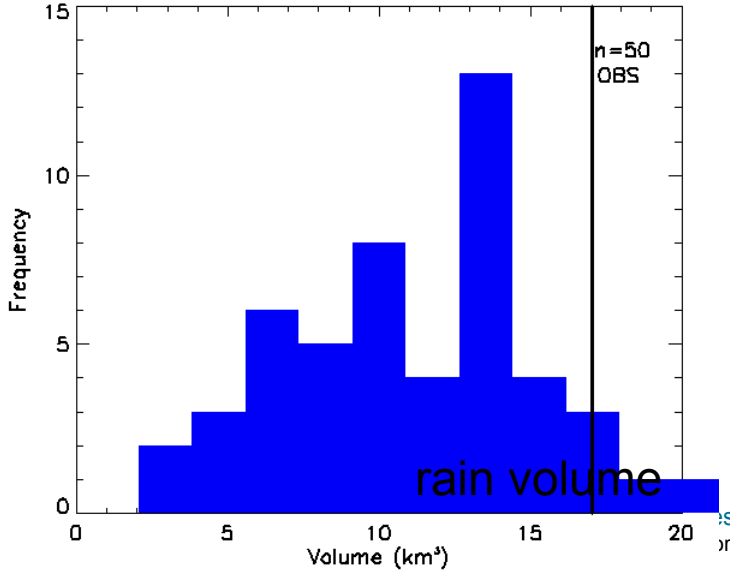
Distributions of feature attributes





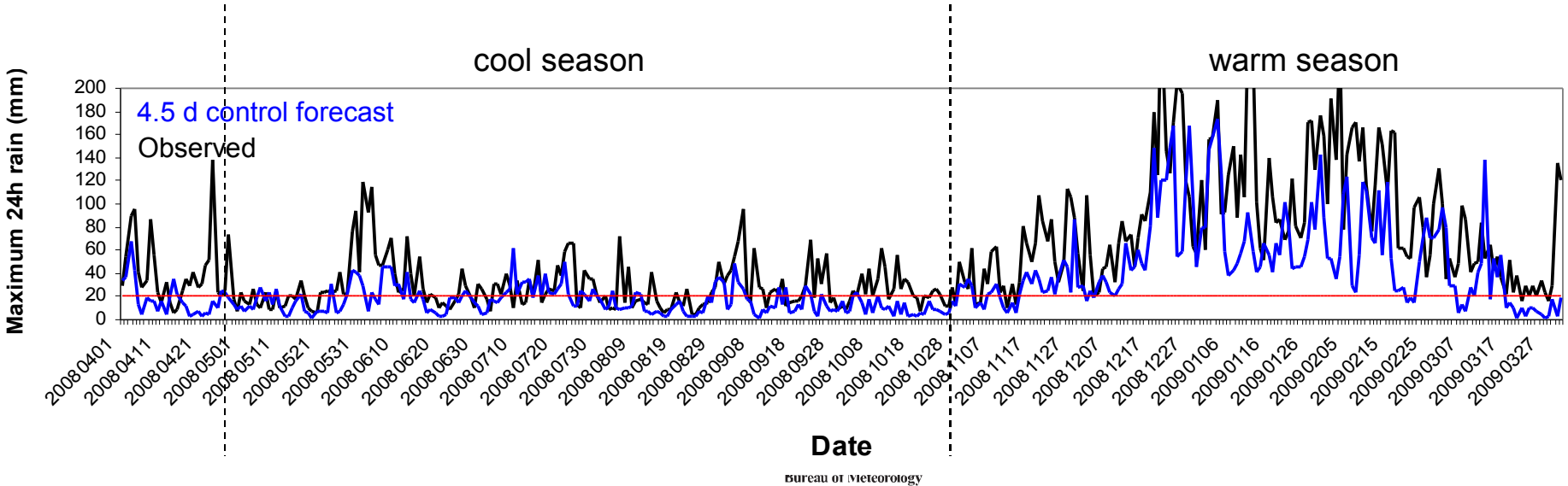






ECMWF EPS heavy rain verification
1 April 2008 – 31 March 2009

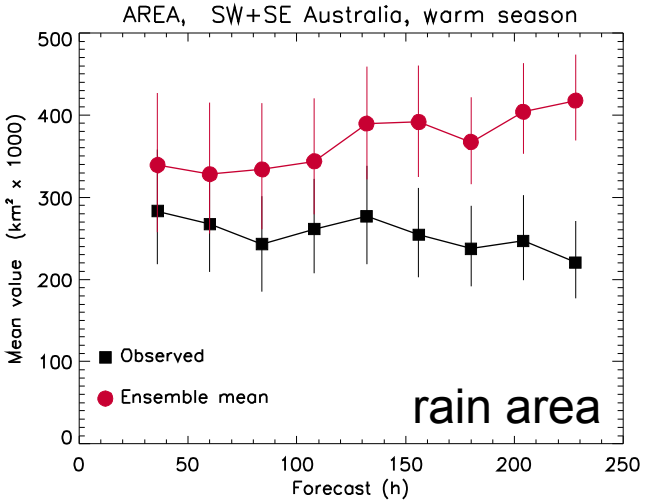
- ECMWF EPS @ T_L399L91 resolution (~0.5º)
- Forecasts of 24h rain accumulation to 9½ days
- Verified against BOM operational daily rain gauge analysis (0.25º)
- 0.5º verification grid
- Verify individual ensemble members' CRAs defined by 5 mm d^{-1} threshold
- Focus on events with observed or forecast maxima of ≥ 20 mm d^{-1}
- Stratify results by season and region

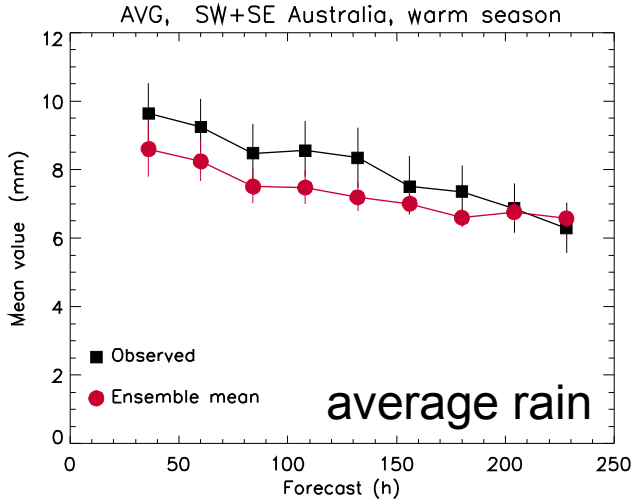


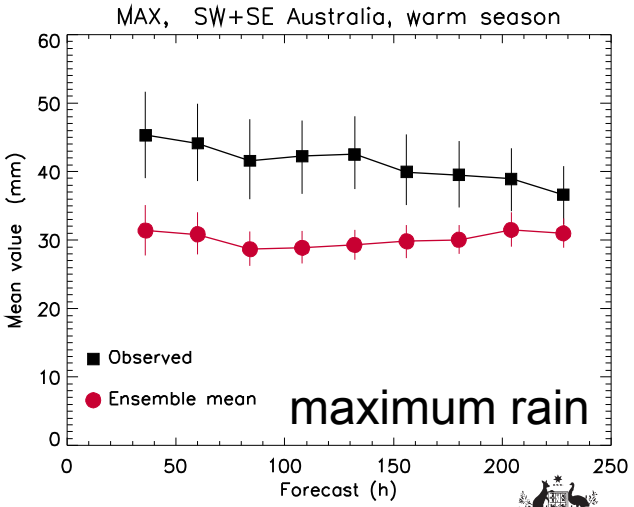
Ensemble mean values

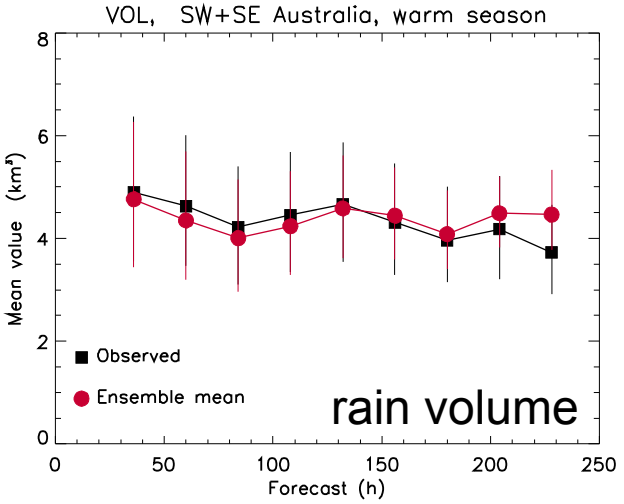
CRAs with max rain ≥ 20 mm d^{-1} Warm season, southern Australia

Mean values for 112 events







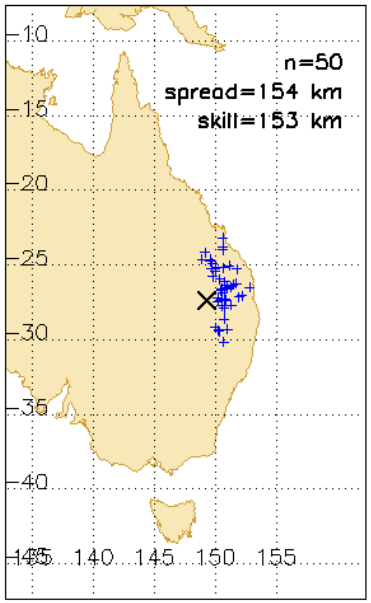


Australian Government
Bureau of Meteorology

The Centre for Australian Weather and Climate Research
A partnership between CSIRO and the Bureau of Meteorology

CSIRO

Mean spread and skill for location forecasts

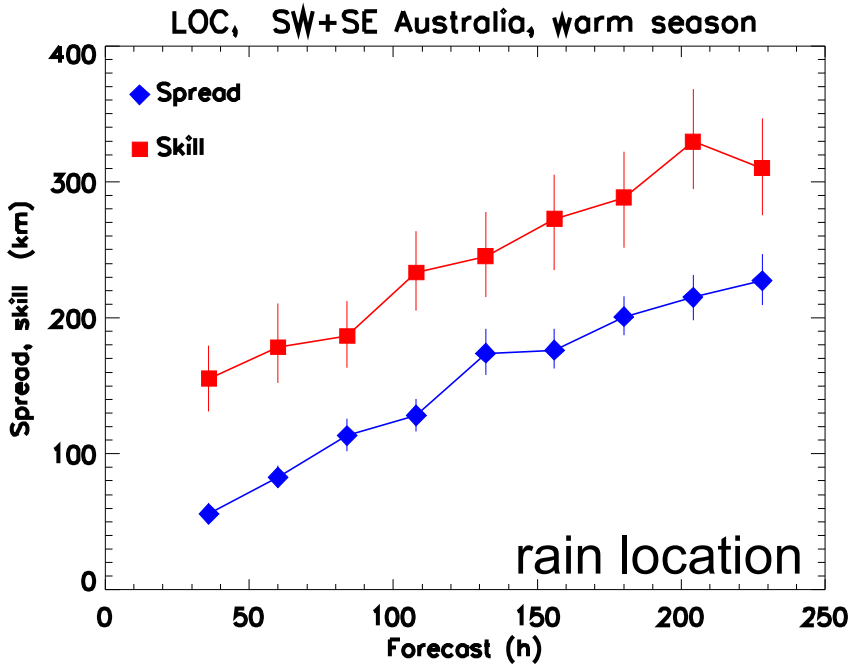


Spread = average distance to ensemble median location

Skill = distance between ensemble median and observed location

CRAs with max rain ≥ 20 mm d^{-1}
Warm season, southern Australia

Mean values for 112 events

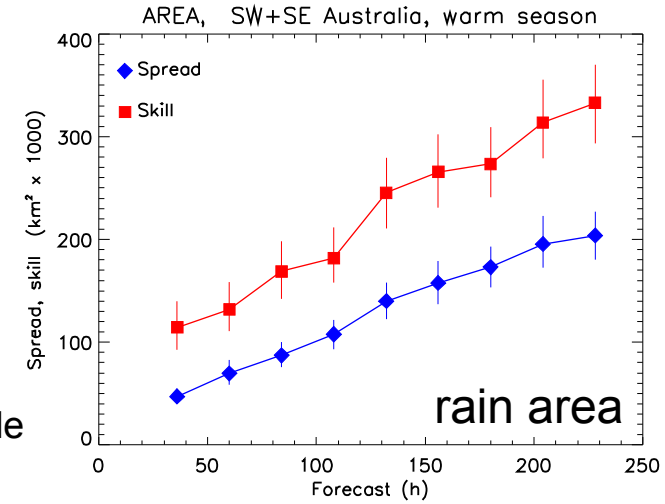


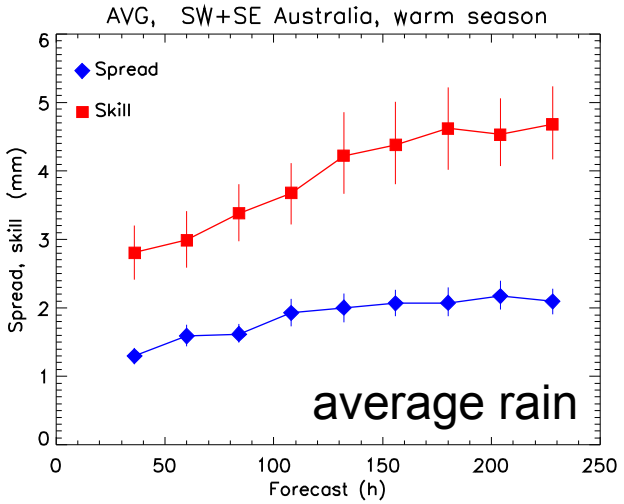
Mean spread and skill

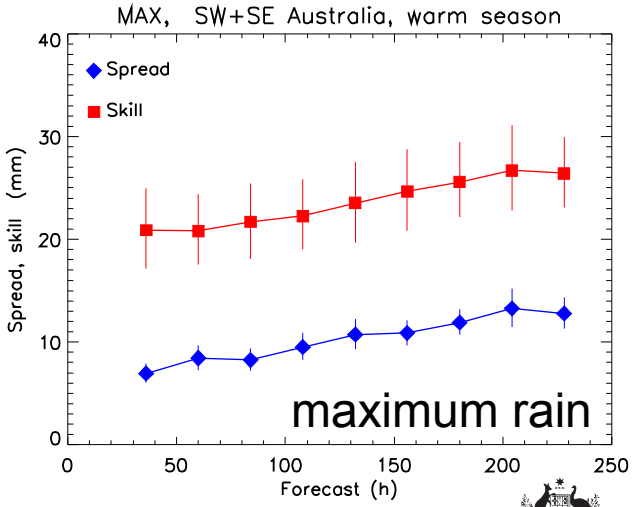
CRAs with max rain ≥ 20 mm d^{-1} Warm season, southern Australia

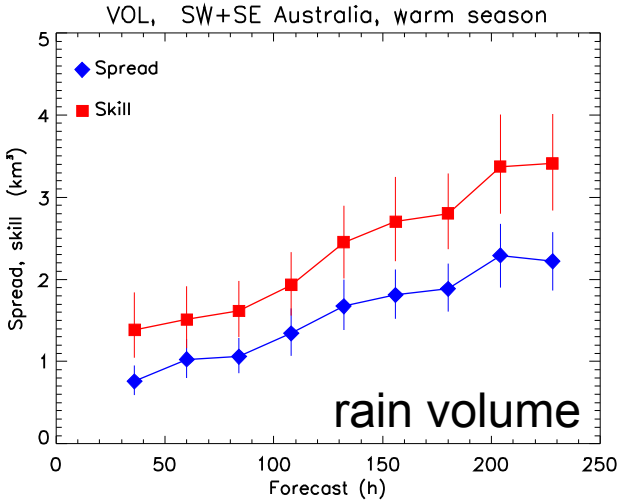
Spread = SD of ensemble attributes

Skill = RMSE of ensemble mean attribute









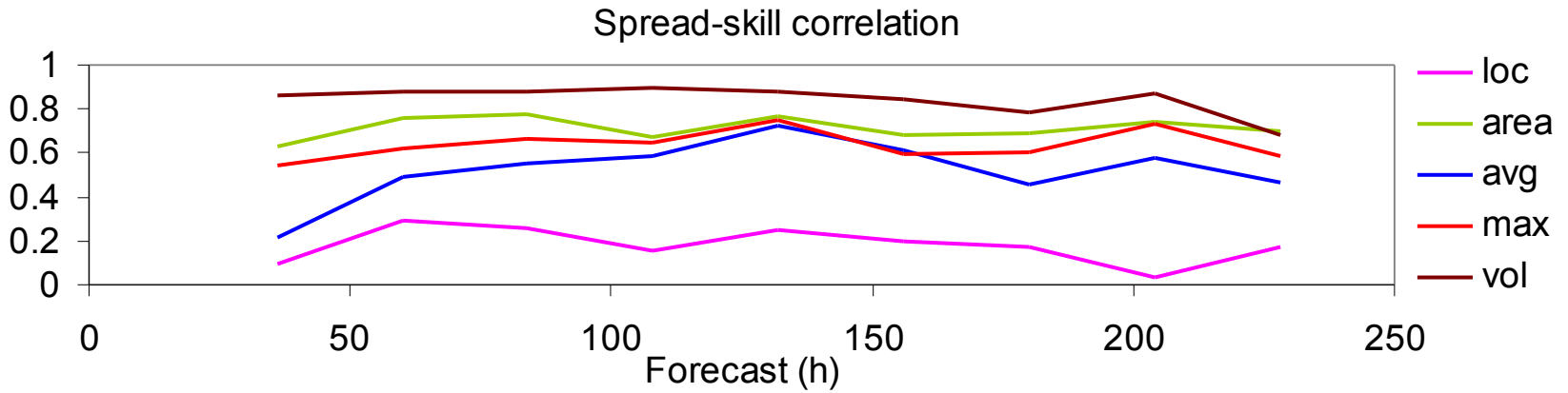
Australian Government
Bureau of Meteorology

The Centre for Australian Weather and Climate Research
A partnership between CSIRO and the Bureau of Meteorology

CSIRO

Feature spread-skill correlations

CRAs with max rain ≥ 20 mm d^{-1} Warm season, southern Australia

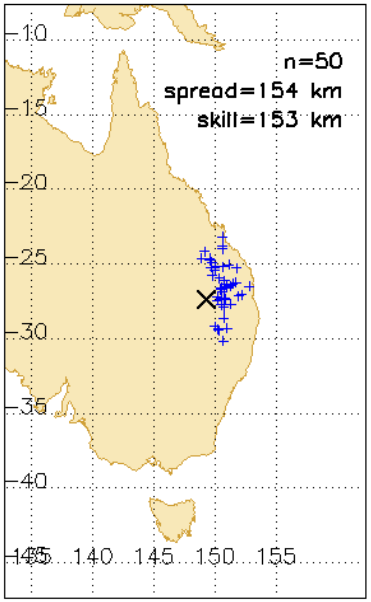


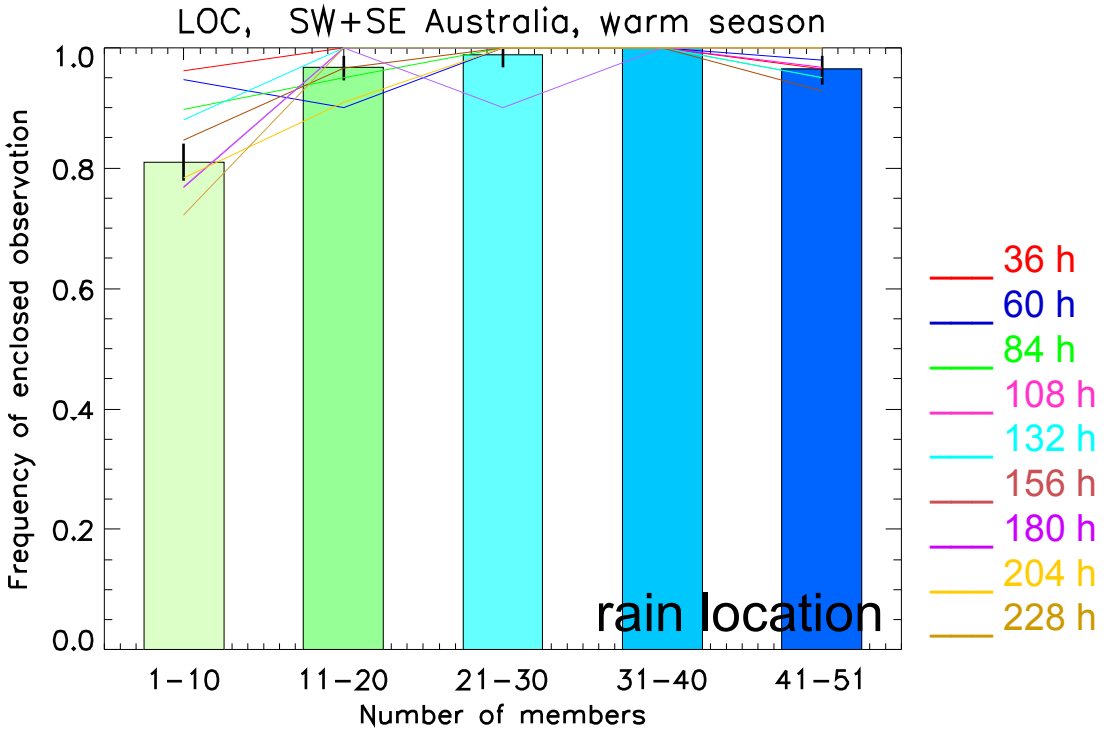
Attribute	object location	object area	object mean rainfall	object maximum rainfall	object rain volume
Mean spread-skill correlation	0.18	0.71	0.52	0.64	0.84

→ Uncertainty in heavy rain system properties can be predicted reasonably well except for location

How often is the observed value within the ensemble distribution?

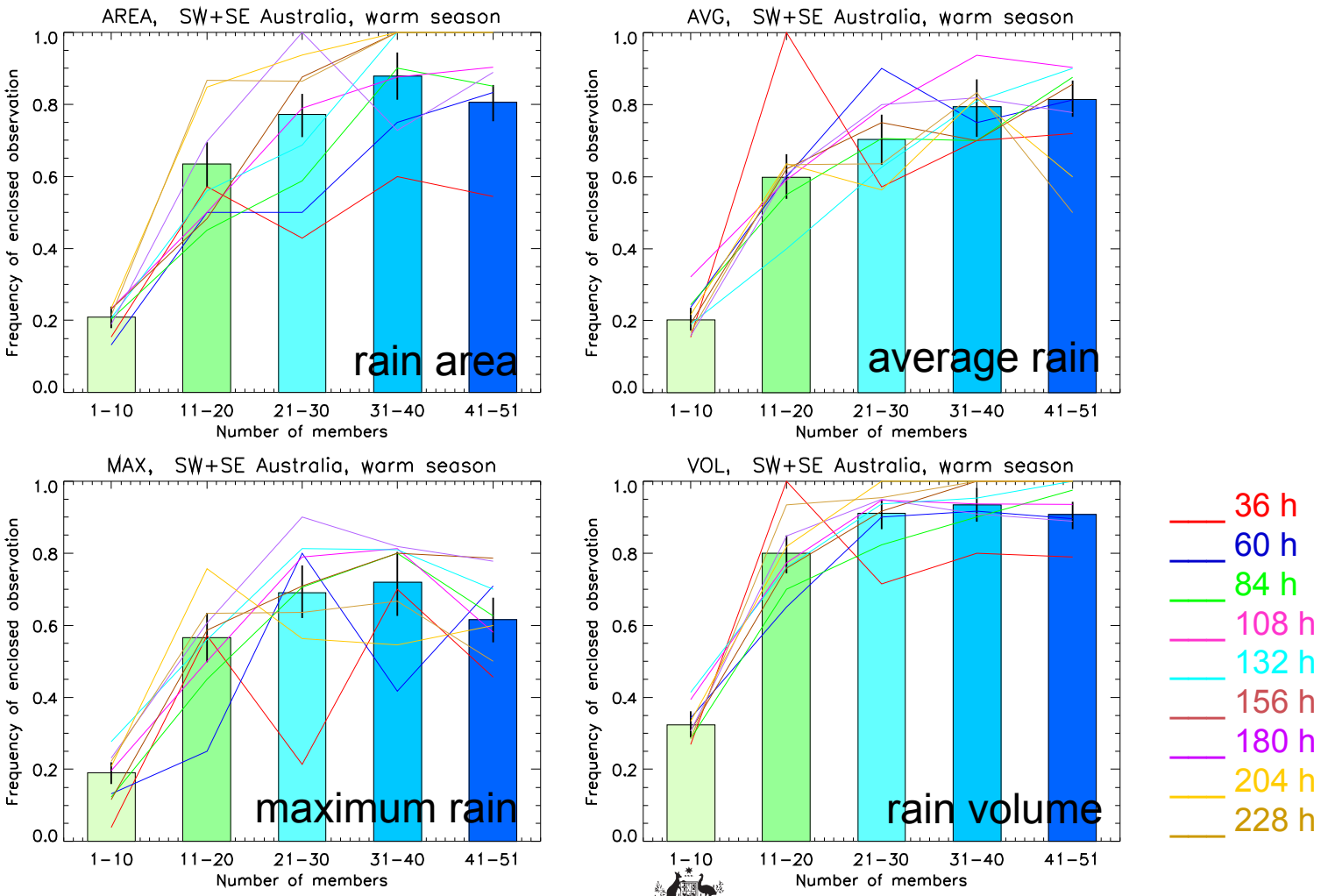
CRAs with max rain ≥ 20 mm d^{-1} Warm season, southern Australia





How often is the observed value within the ensemble distribution?

CRAs with max rain ≥ 20 mm d^{-1} Warm season, southern Australia



Australian Government
Bureau of Meteorology

The Centre for Australian Weather and Climate Research
A partnership between CSIRO and the Bureau of Meteorology

CSIRO

Probabilistic verification

- How accurate is the **distribution of feature attributes** for the ensemble members?
- What thresholds to use? (have already selected for heavy rain events)

Attribute	Median value
area of feature	2×10^5 km^2
mean rainfall	10 mm d^{-1}
maximum rainfall	40 mm d^{-1}
rain volume	2.5 km^3

Forecast probability = fraction of ensemble members ≥ threshold

Australian Government
Bureau of Meteorology

The Centre for Australian Weather and Climate Research
A partnership between CSIRO and the Bureau of Meteorology

CSIRO

Does the number of ensemble members predicting an event have meaning?

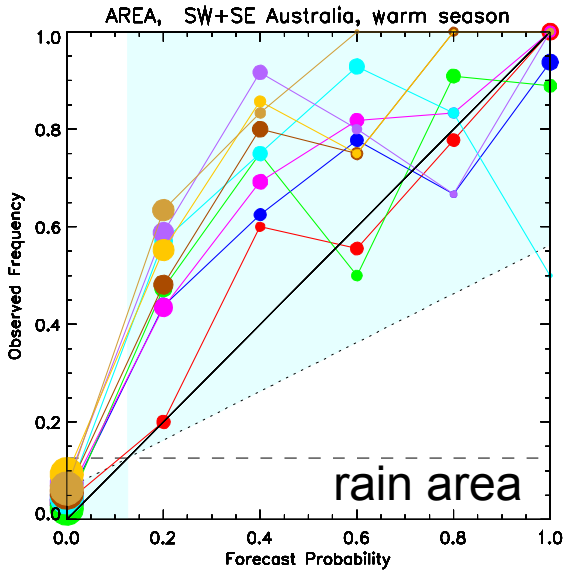
fraction of Forecast =

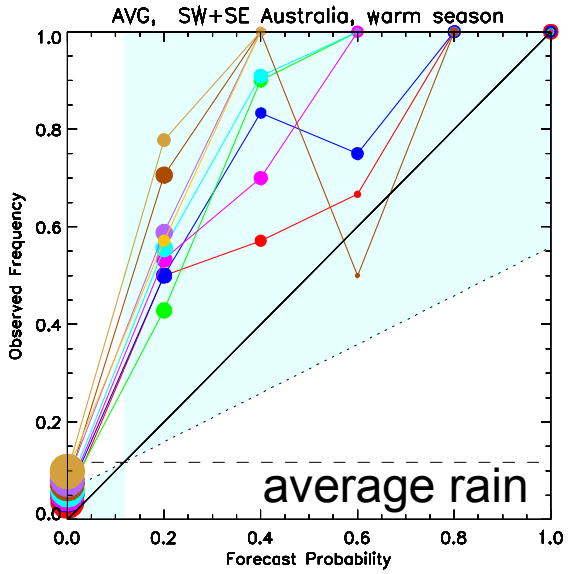
ensemble probability

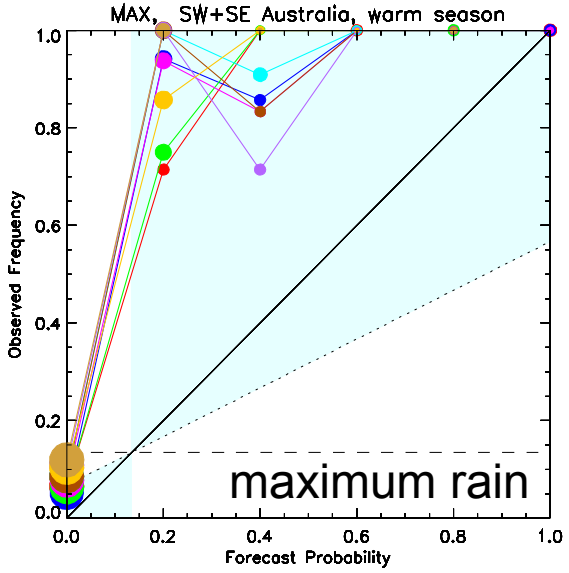
members

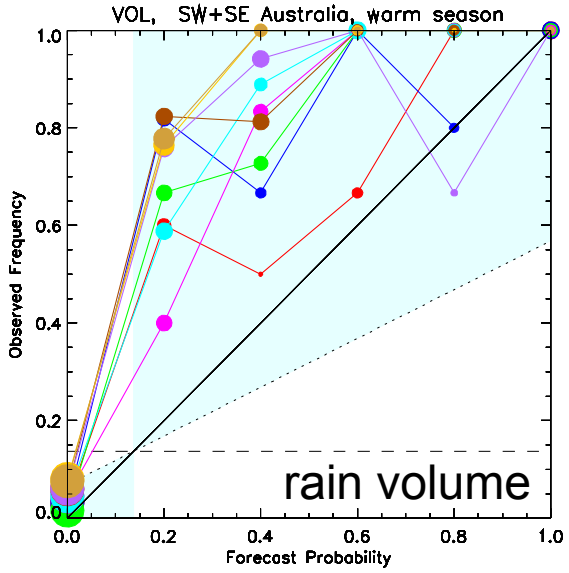
≥ threshold

36 h
60 h
84 h
108 h
132 h
156 h
180 h
204 h
228 h



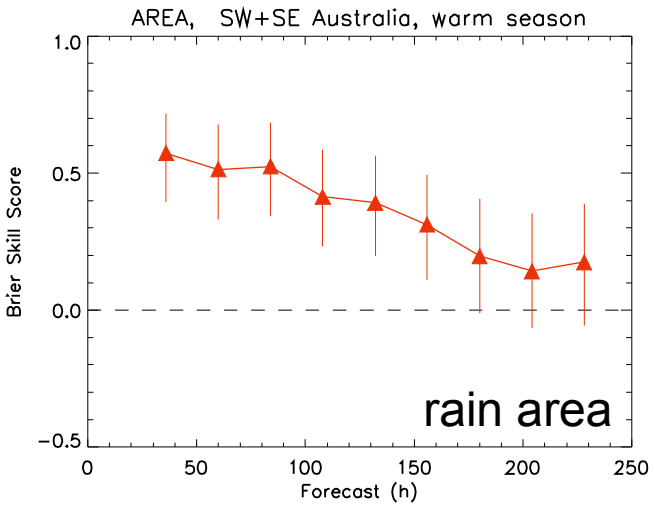


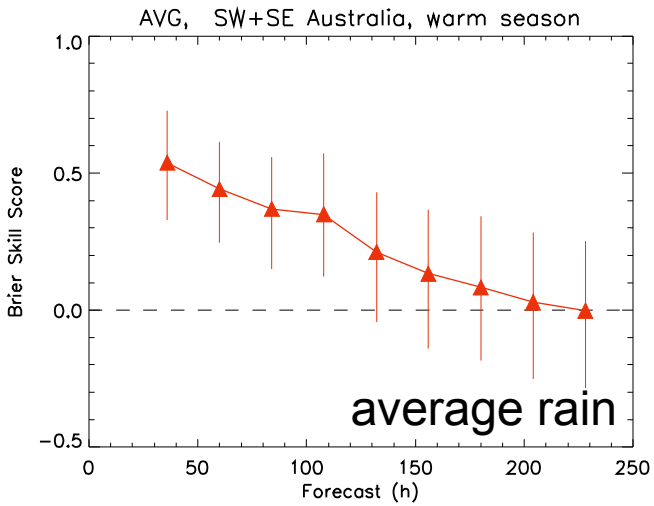


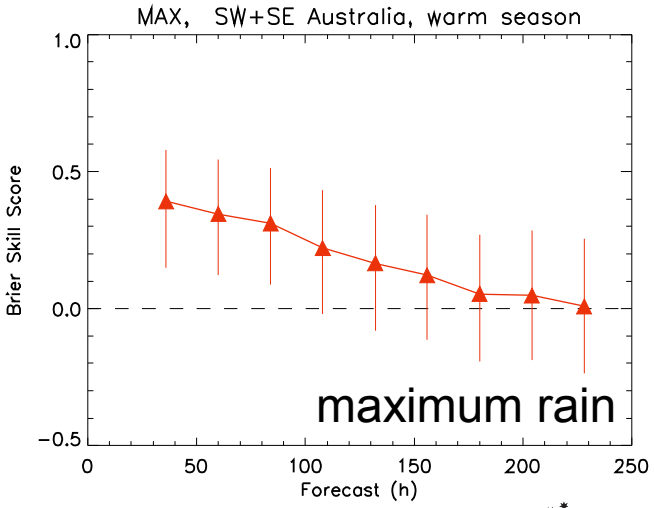


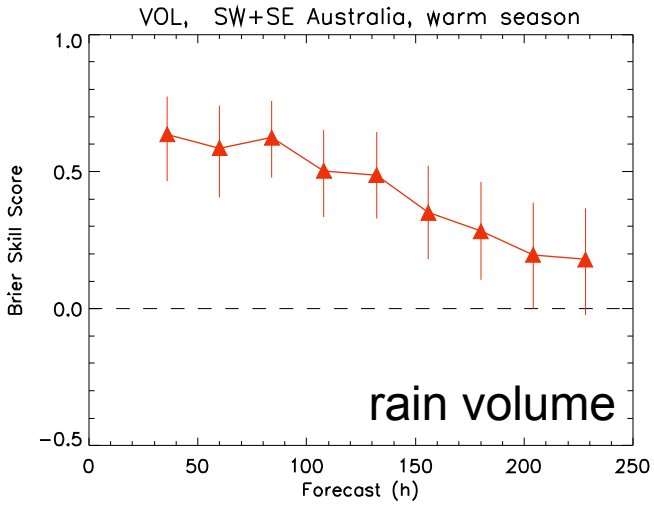
Does the number of ensemble members predicting an event have meaning?

Brier skill score with respect to sample climatology of observed events









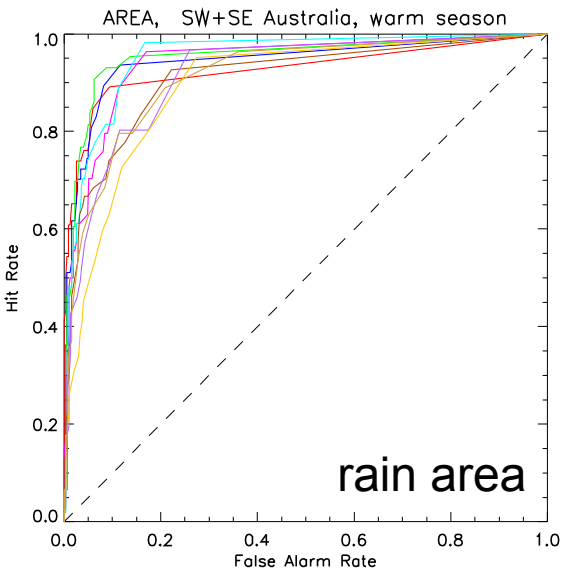
Australian Government
Bureau of Meteorology

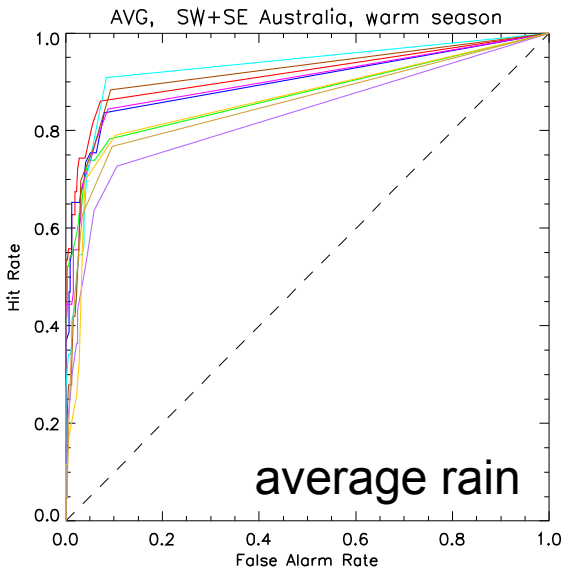
The Centre for Australian Weather and Climate Research
A partnership between CSIRO and the Bureau of Meteorology

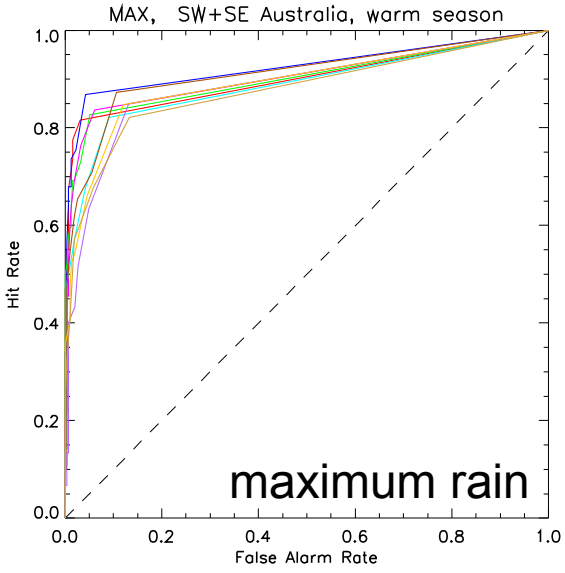
CSIRO

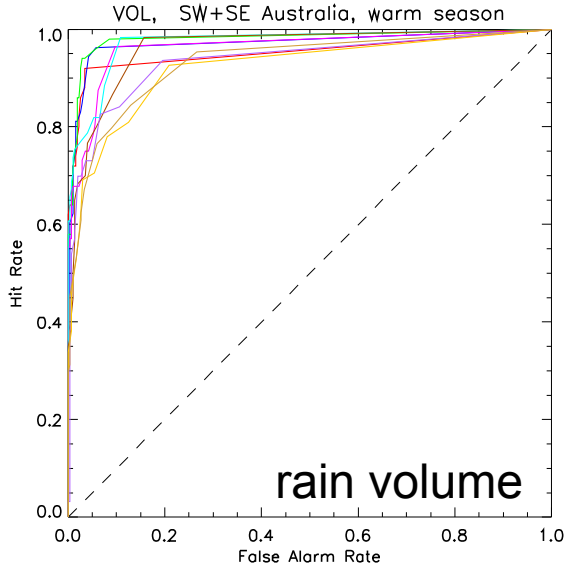
Does the number of ensemble members predicting an event have meaning?

ROC varies the number of ensemble members exceeding the threshold required for a "yes event" forecast









36 h
60 h
84 h
108 h
132 h
156 h
180 h
204 h
228 h

Early results

- **Feature-based verification looks very promising for evaluation of heavy rain events and other weather features**
- The forecast feature attributes (location, rain area, mean & max rain, rain volume) showed some mean biases
- The ensemble spread is less than the RMSE of the ensemble mean for all feature attributes
- Spread-skill correlation is high (except for location)
  → **spread can be used to predict uncertainty**
- The more members predicting an event, the greater the chance of its observed properties being enveloped by the ensemble
- Probability forecasts for events were under-confident
- Brier skill score consistently positive
- ROC shows high discrimination ability → calibration should help!

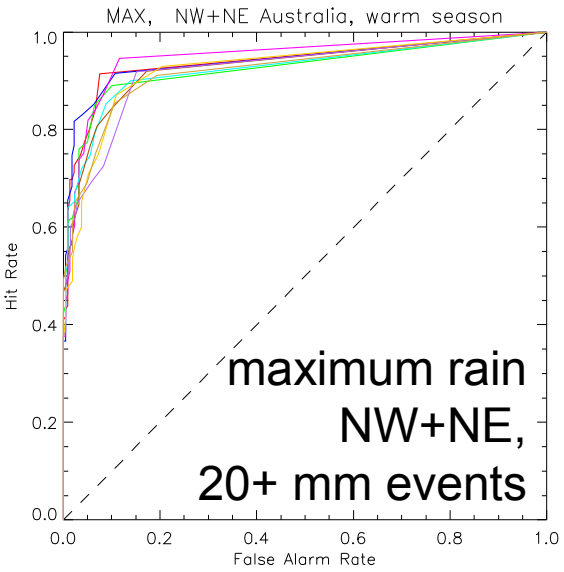
Thank you!

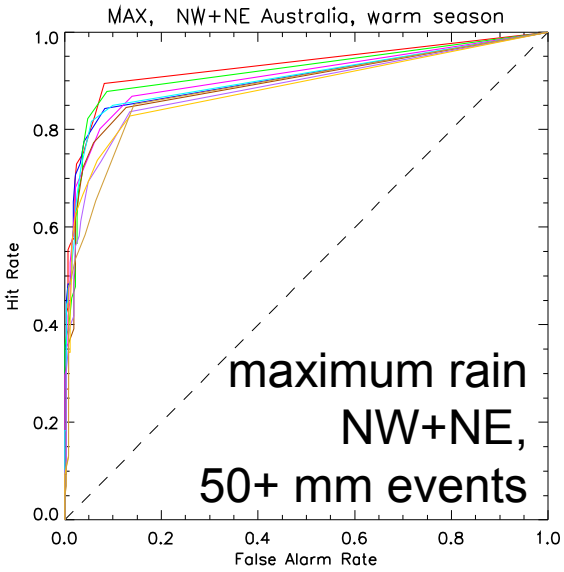
How do results depend on the magnitude of the event?

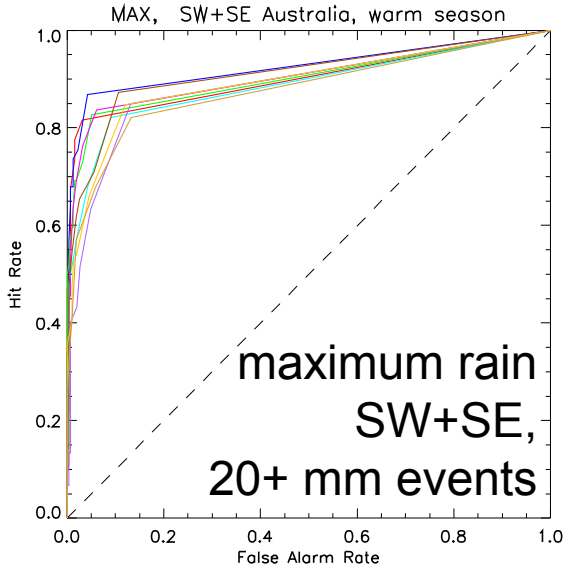
Thresholds for verifying maximum rain:

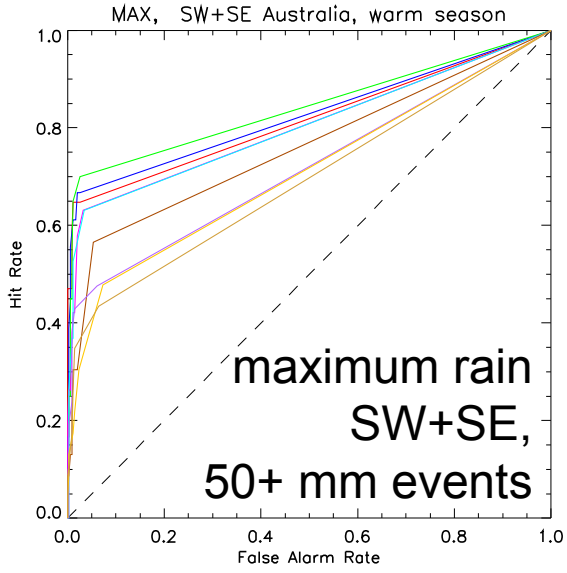
20+ mm events: 40 mm

50+ mm events: 75 mm









36 h
60 h
84 h
108 h
132 h
156 h
180 h
204 h
228 h