# Verification of Sub-seasonal to Seasonal Predictions

***Caio Coelho***
***INPE/CPTEC, Brazil***
***caio.coelho@cptec.inpe.br***

*Acknowledgments: Arun Kumar, Alberto Arribas, Barbara Brown, Beth Ebert, David Stephenson, Debbie Hudson, Laura Ferranti, Matthew Wheeler, Simon Mason and Yuhei Takaya*

**Plan of talk**

1) Brief review of forecast goodness
2) Attributes based forecast quality assessment: examples o
3) Final remarks

7th International Verification Methods Workshop
*Tutorial on forecast verification methods*
*Berlin, Germany, 3-6 May 2017*

JWGFVR

# What is a good forecast?

**Good forecasts have:**

- QUALITY
- VALUE/UTILITY
- CONSISTENCY


**Attributes of quality:**

- Association
- Accuracy
- Discrimination
- Reliability
- Resolution

…

→ No single score can be used to summarize a set of forecasts

A. H. Murphy 1993
"What is a good forecast ?
An essay on the nature of goodness in weather forecasting"
Weather and Forecasting, 8, 281-293.

# Some definitions

- Quality: Measure of correspondence between forecasts and observations using mathematical relationship (deterministic and probabilistic scores)
- Value: Measure of benefit achieved (or loss incurred) through the use of forecasts
- Consistency: Correspondence between a forecast and the forecasters belief. If consistent, the forecast must communicate what the forecaster thinks will happen, and correctly indicate the associated level of uncertainty

# S2S forecast quality assessment

# 1. Attributes of deterministic forecasts (ensemble mean)

# Association

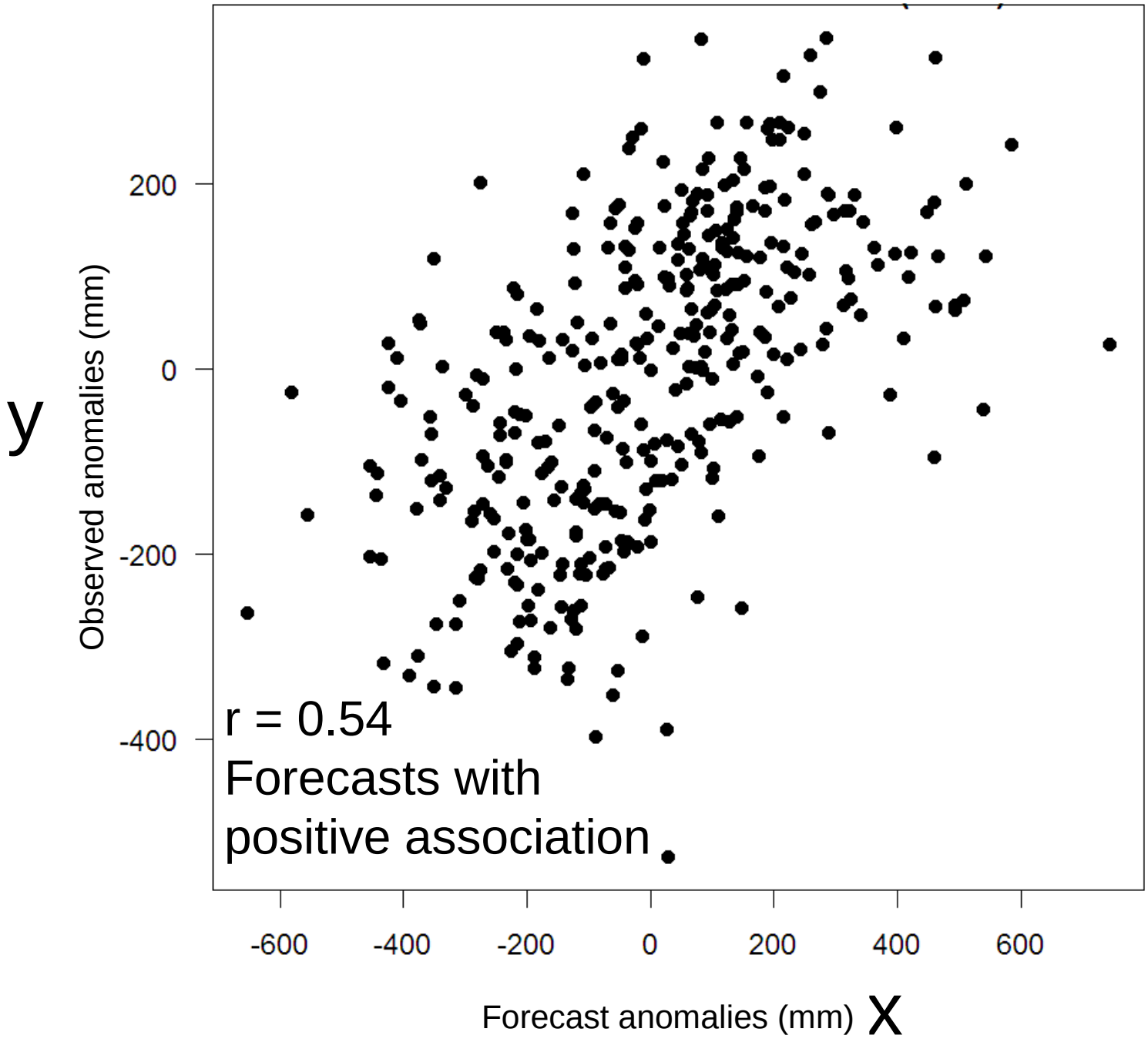
- Overall strength of the relationship between the forecasts and observations
- Linear association is often measured using the product moment **correlation coefficient**

$$r = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2}}$$

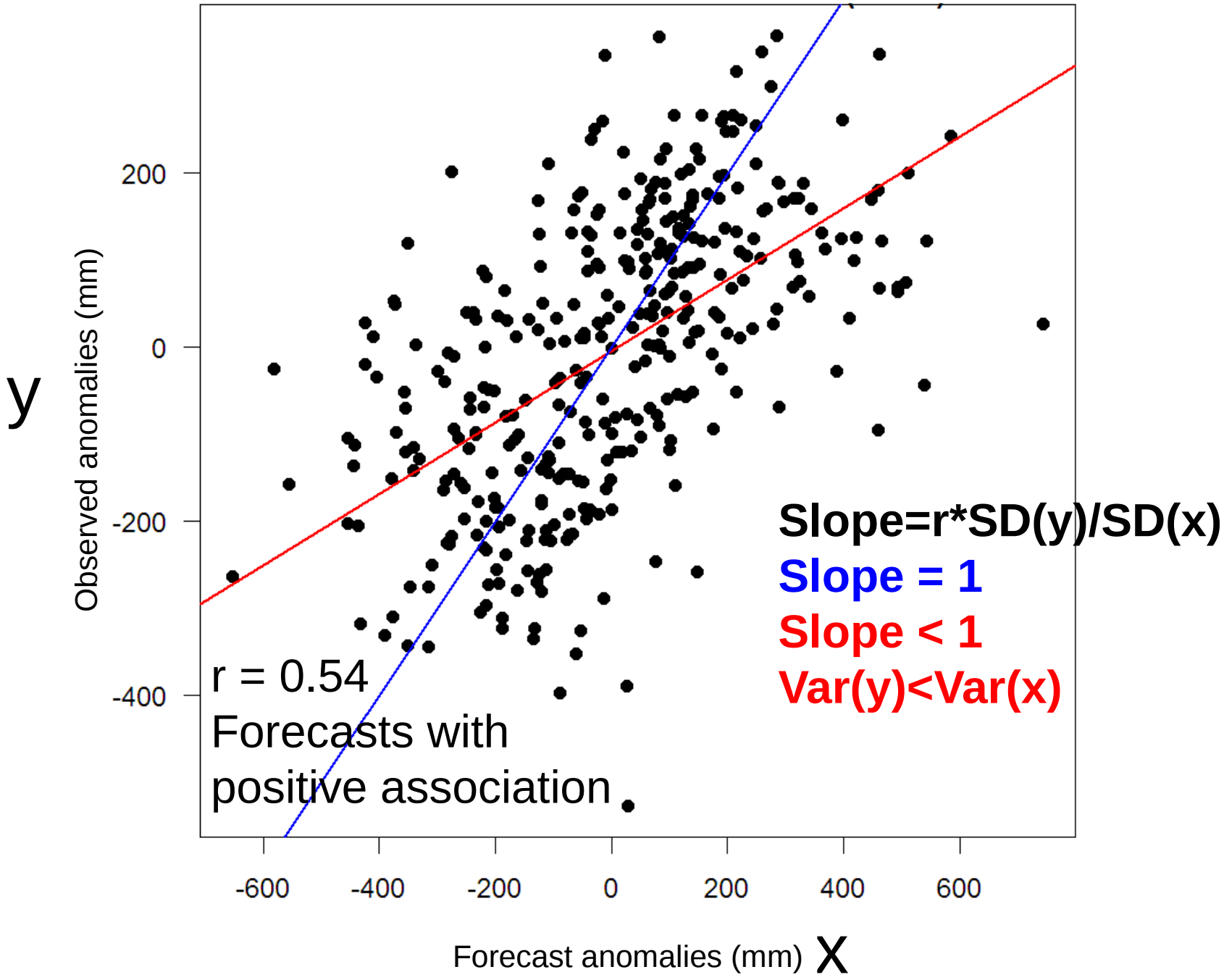
*x: forecast* *y: observation*

*n: number of (x,y) pairs*

# Relationship between past forecast and past obs. anomalies



# Relationship between past forecast and past obs. anomalies



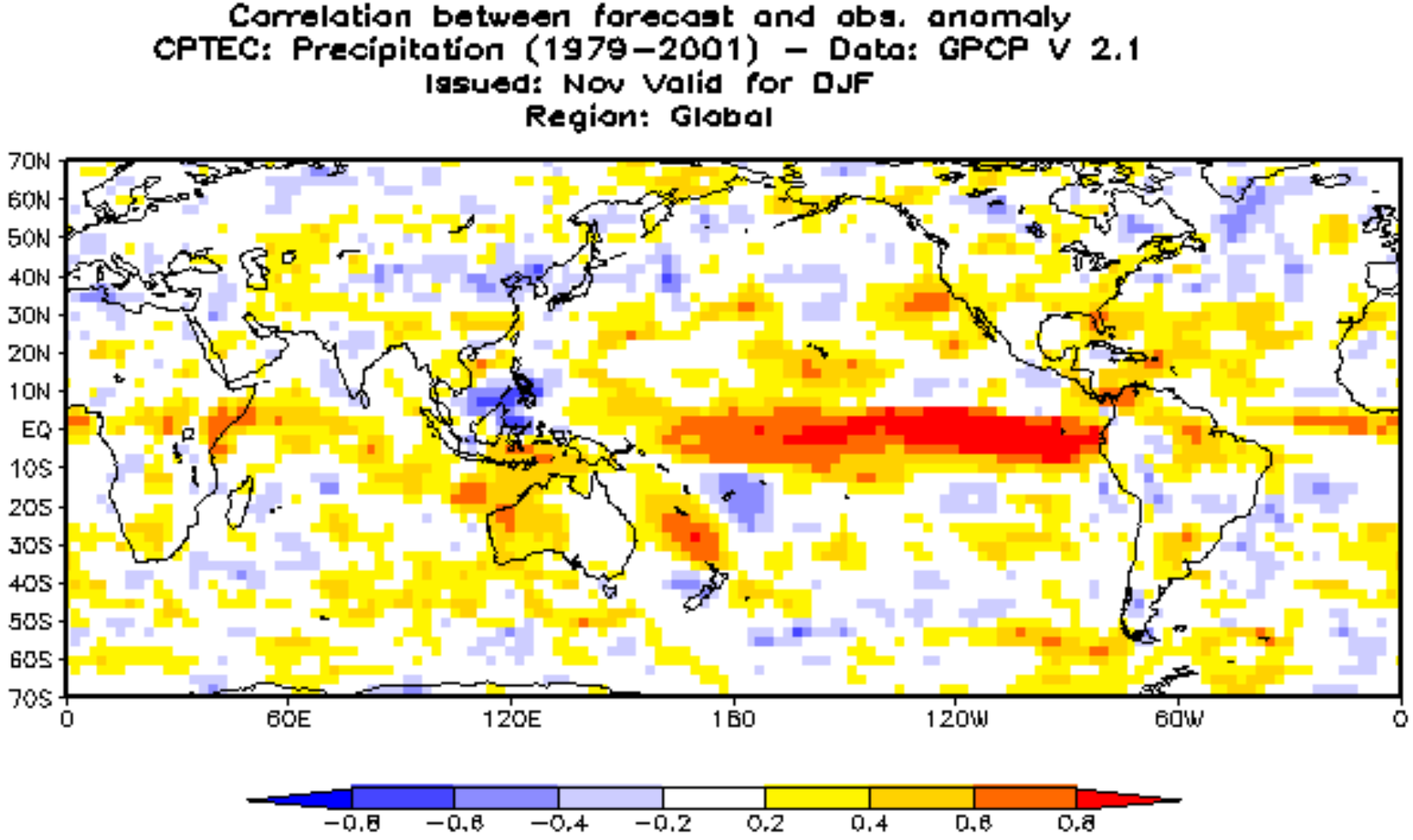
# Accuracy

- Average distance between forecasts and observations
- Simplest measure is the **Mean Error (Bias)**

$$ME = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - y_i\right)$$

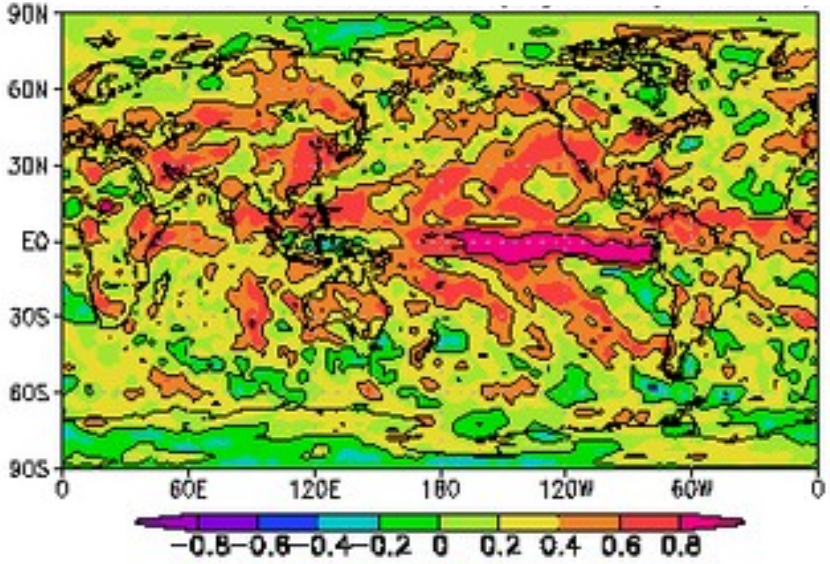
*x: forecast  y: observation  n: number of (x,y) pairs*

# Seasonal forecast example: 1-month lead precip. fcsts for DJF

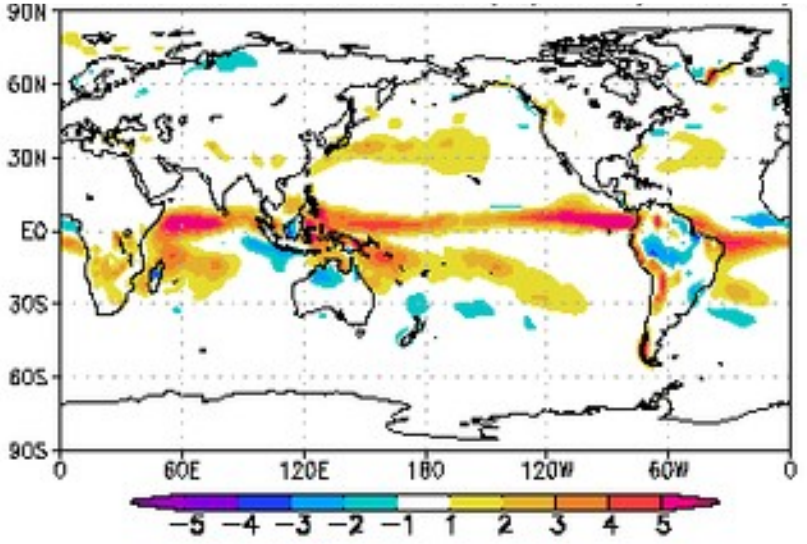


# Monthly forecast example:
# 0-day lead precip. fcsts for next 30 days

ACC (against GPCP v2 monthly)
Day 1-30 mean
I.C. : Dec.-Feb. 1981-2010



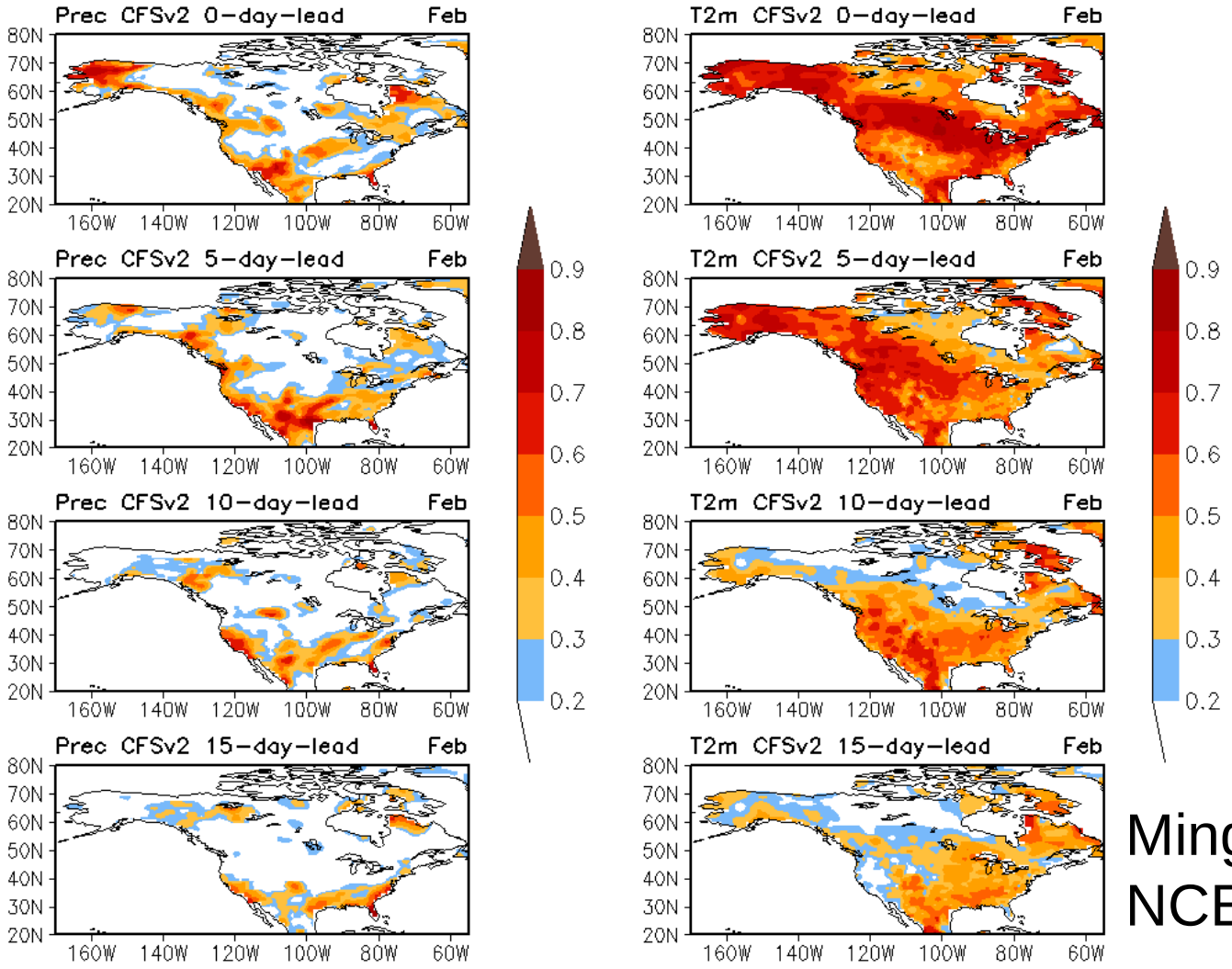

Bias (against GPCP v2 monthly)
Day 1-30 mean
I.C. : Dec.-Feb. 1981-2010




Yuhei Takaya, JMA

# Monthly forecast example:
# 0, 5, 10 and 15-day lead fcsts for Feb

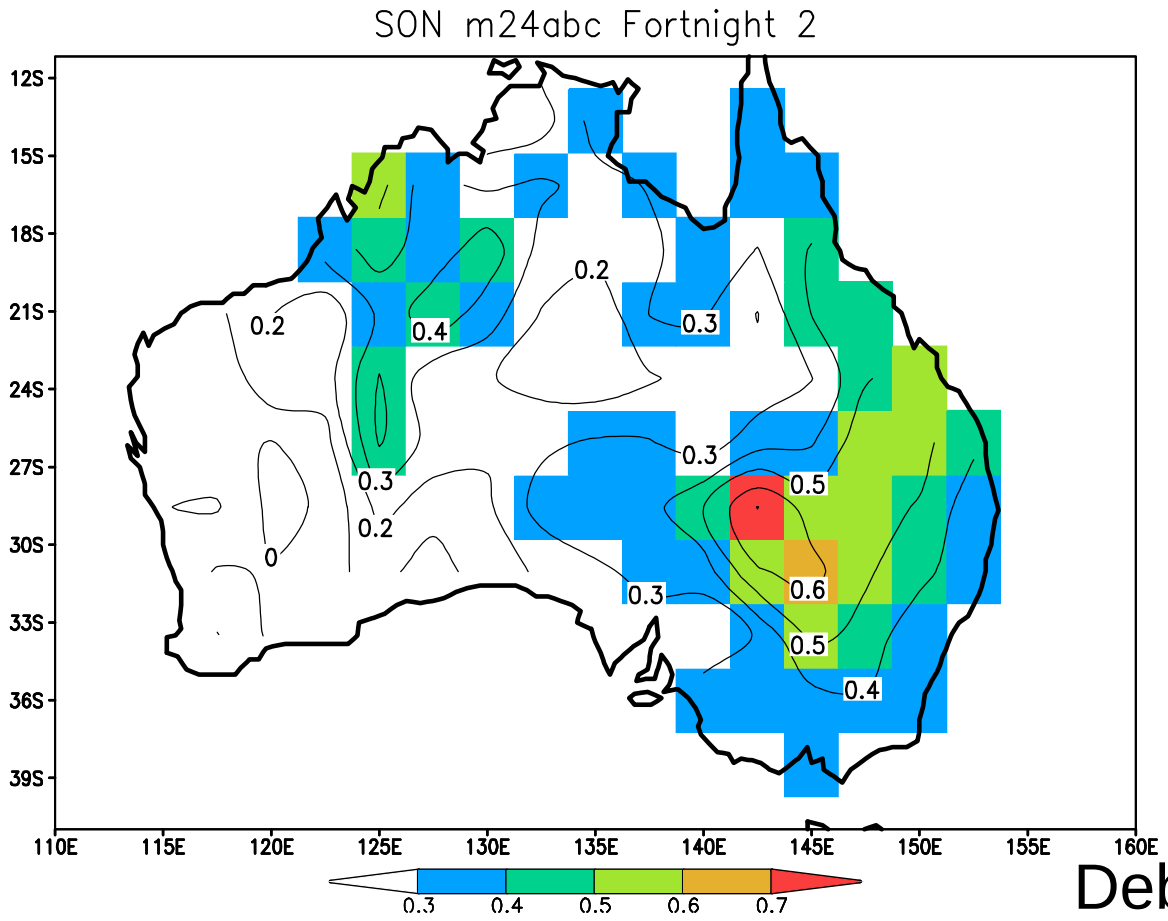
**Precipitation** CFSv2 Correlation Feb (1982–2009) **2m Temperature**




Mingyue Chen
NCEP/NOAA

# Two weeks forecast example: ½ month lead precip. fcsts

**Correlation between forecast and observed precipitation anomalies**
**Fortnight 2: Sep, Oct, Nov forecast start months. Hindcasts: 1980-2006**



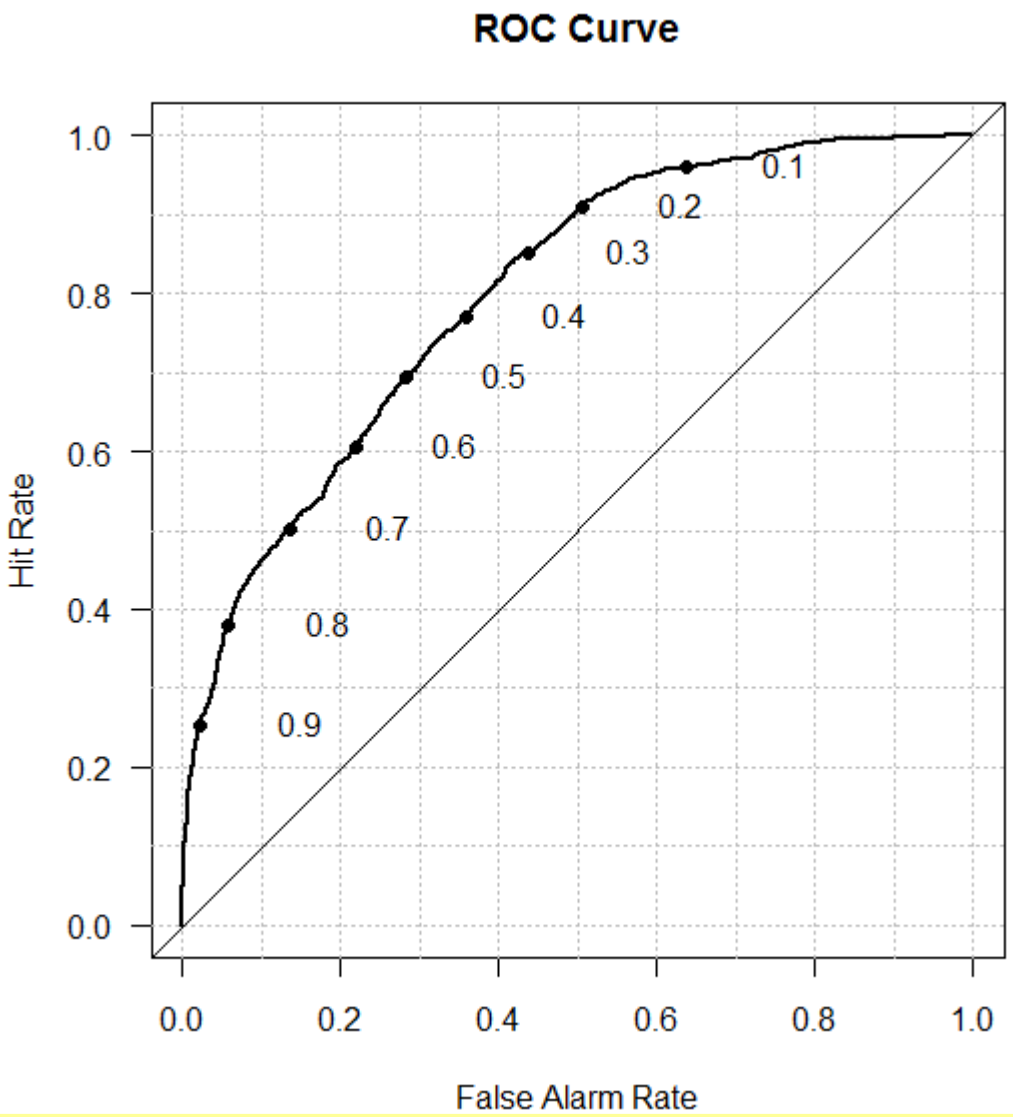

Debbie Hudson
BOM, Australia

# S2S forecast quality assessment

# 2. Attributes of probabilistic forecasts (derived from ensemble members)

# Discrimination

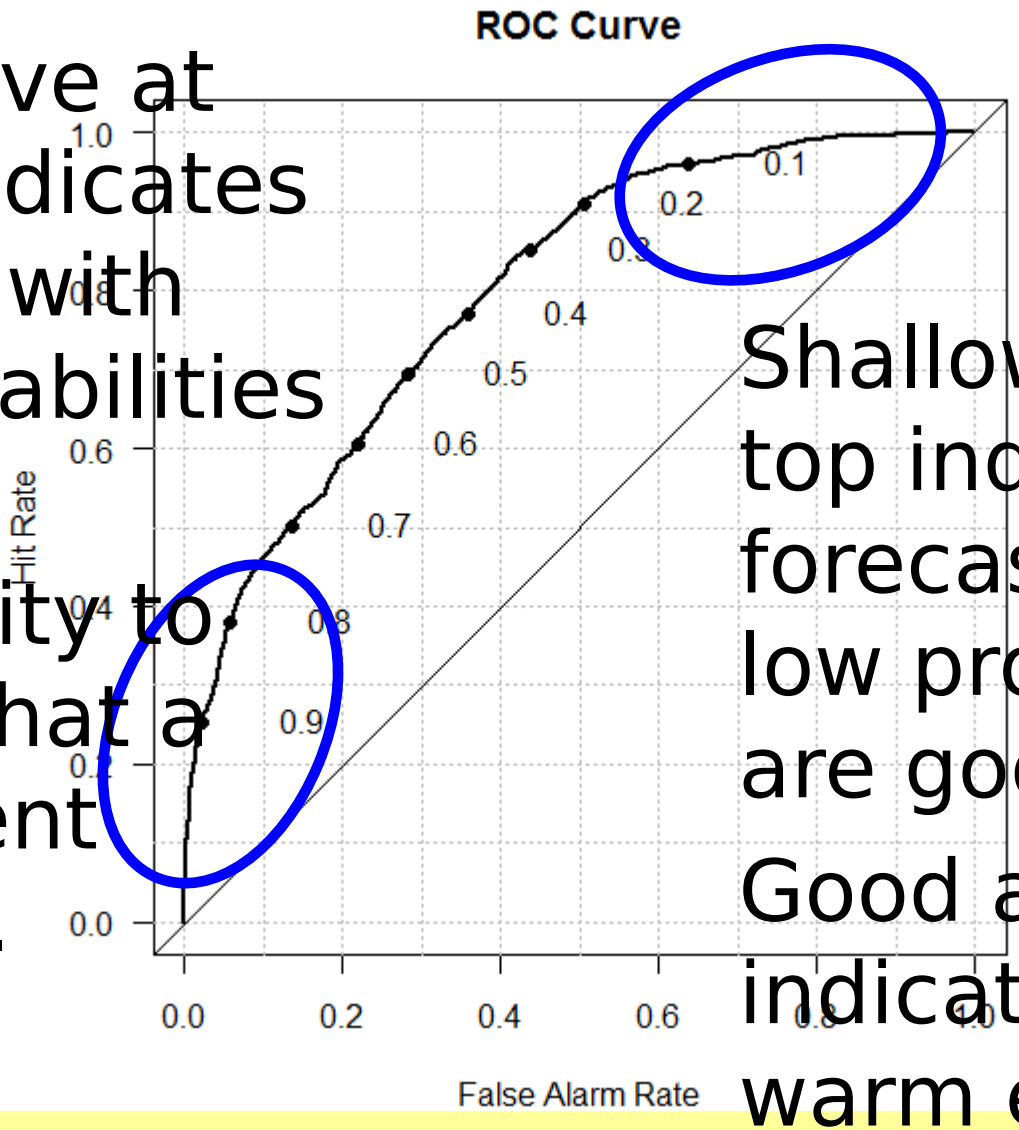
- Conditioning of forecasts on observed outcomes
- Addresses the question: Does the forecast differ given different observed outcomes? Or, can the forecasts distinguish an event from a non-event?
- If the forecast is the same regardless of the outcome, the forecasts cannot discriminate an *event* from a *non-event*
- Forecasts with no discrimination ability are useless because the forecasts are the same regardless of what happens




- The ROC curve is constructed by calculating the hit and
  for various probability thresholds
- Area under ROC curve (A) is a measure of discrimination
  successfully discriminating a warm (SST>0) from a cold (

Steep curve at bottom indicates forecasts with high probabilities are good.

Good ability to indicate that a warm event will occur.

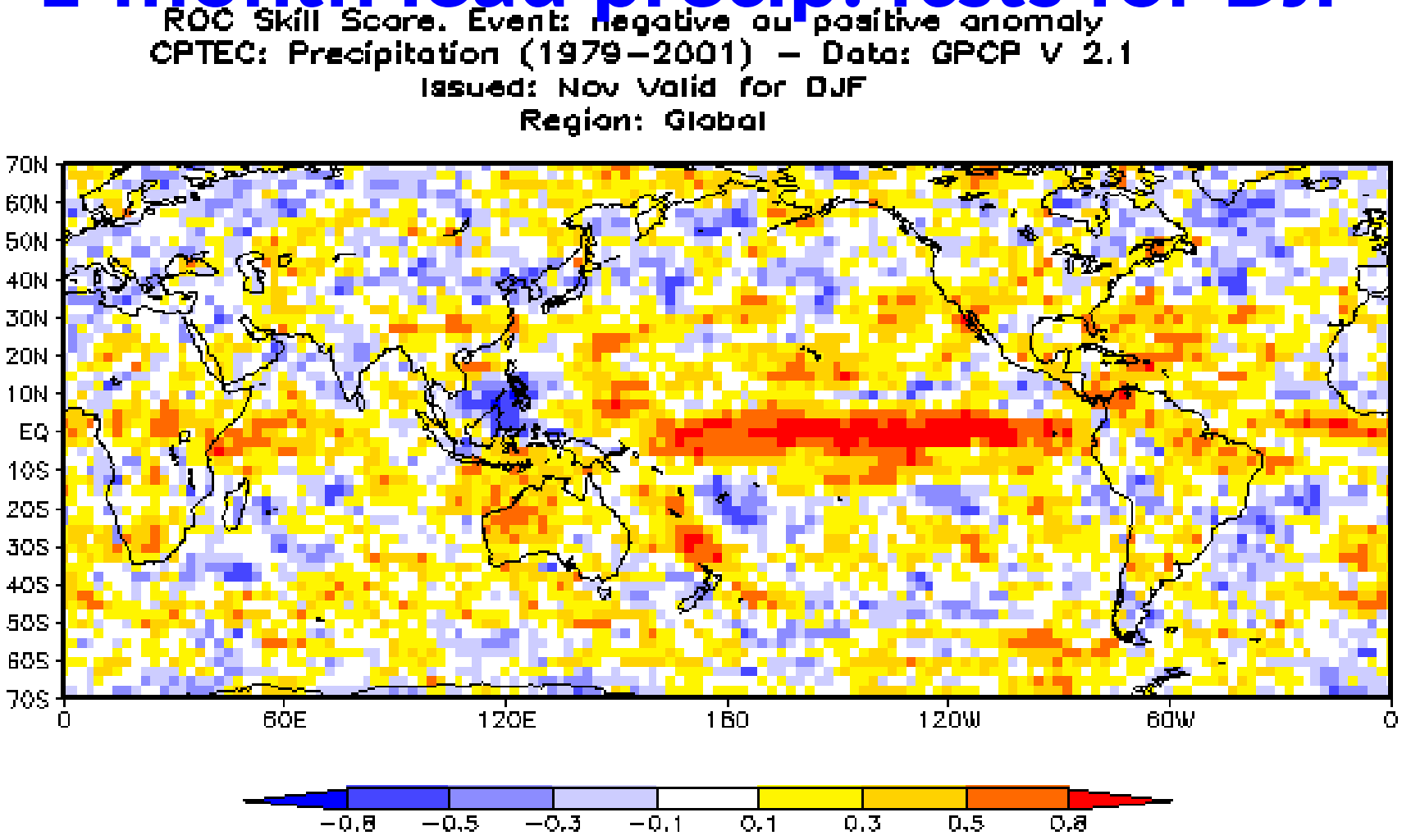



Shallow curve at top indicates forecasts with low probabilities are good.

Good ability to indicate that a warm event will not occur.

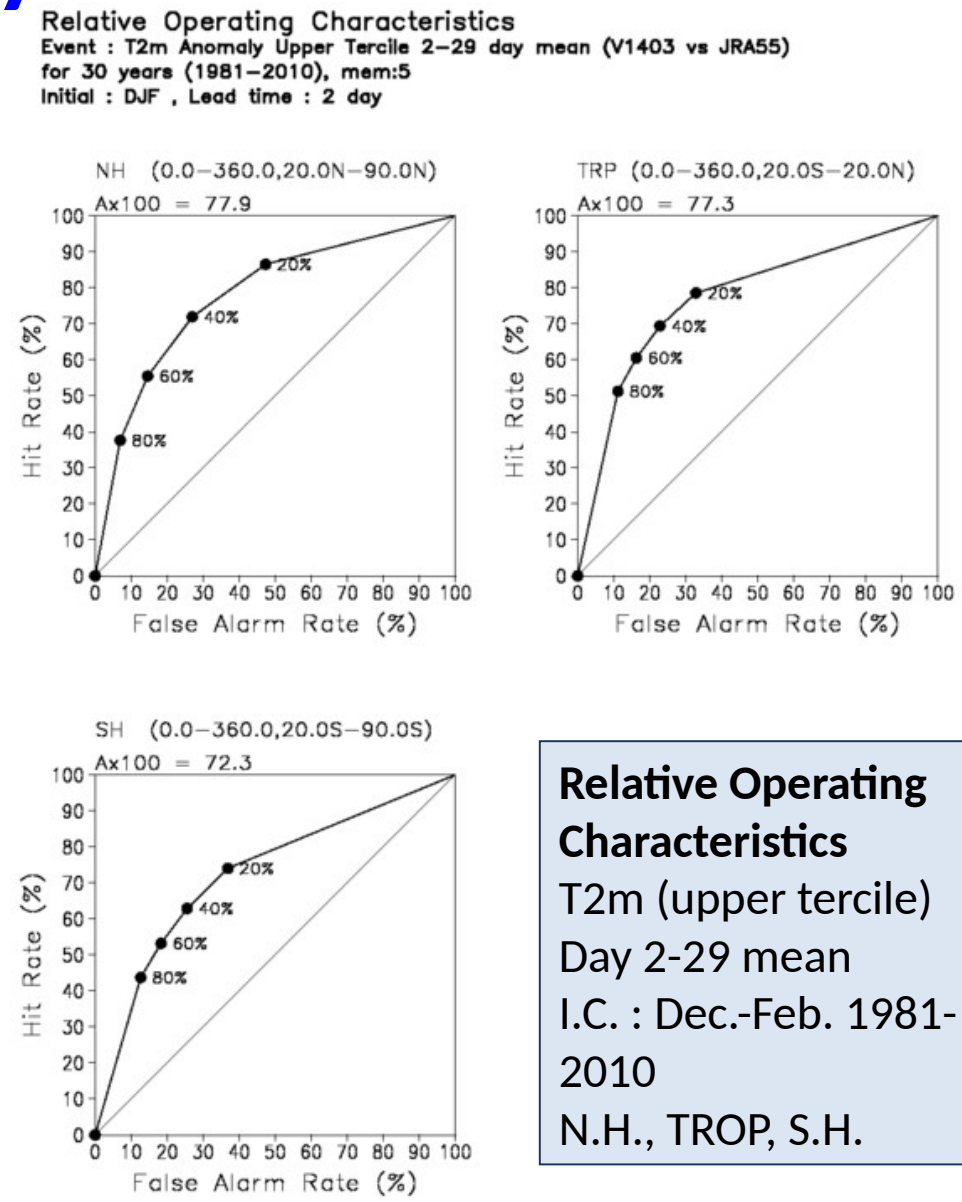
- The ROC curve is constructed by calculating the hit and for various probability thresholds
- Area under ROC curve (A) is a measure of discrimination successfully discriminating a warm (SST>0) from a cold (

# Seasonal forecast example:
# 1-month lead precip. fcsts for DJF




ROC Skill Score = 2 A - 1

# Monthly forecast example:
# 1-day lead 2mT fcsts for day 2-29 mean




**Relative Operating Characteristics**
T2m (upper tercile)
Day 2-29 mean
I.C. : Dec.-Feb. 1981-2010
N.H., TROP, S.H.

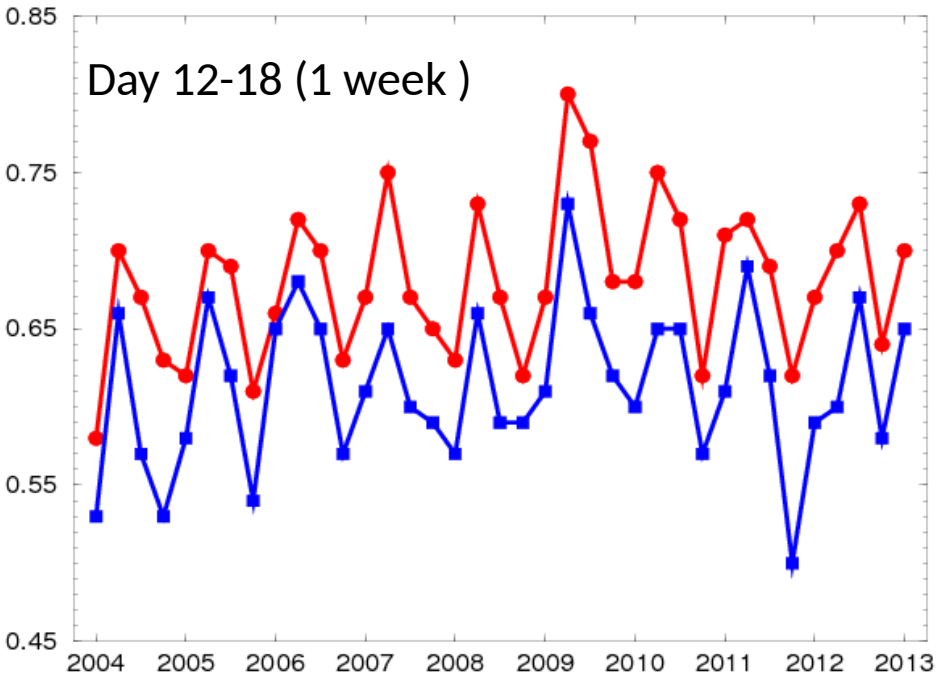
Yuhei Takaya, JMA

# One to two weeks forecast example: Northern extratropics

ROC score: 2-metre temperature in the upper tercile

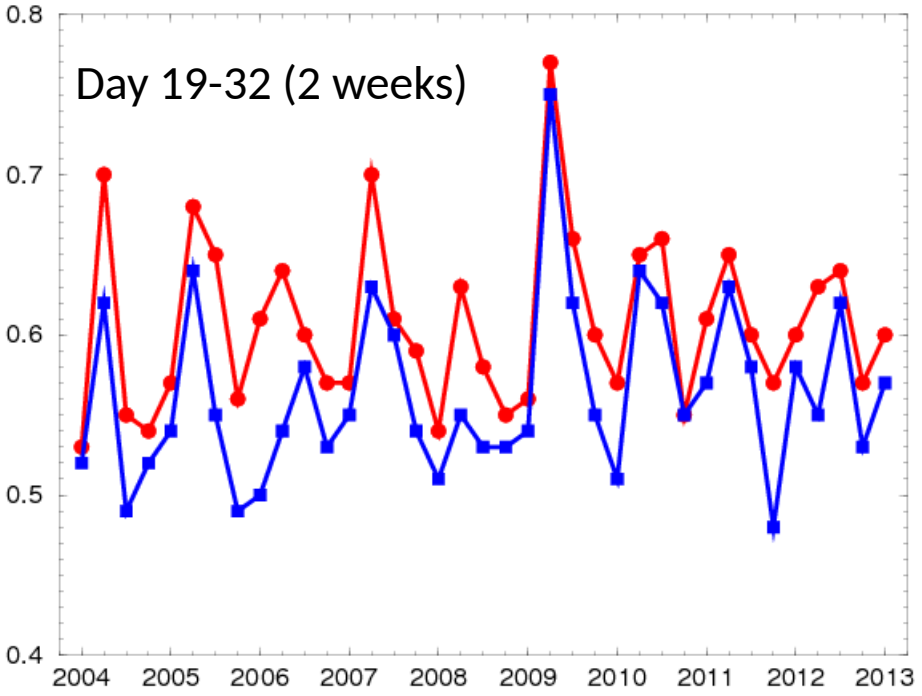
Monthly Forecast

Persistence of day 5-11




Monthly Forecast

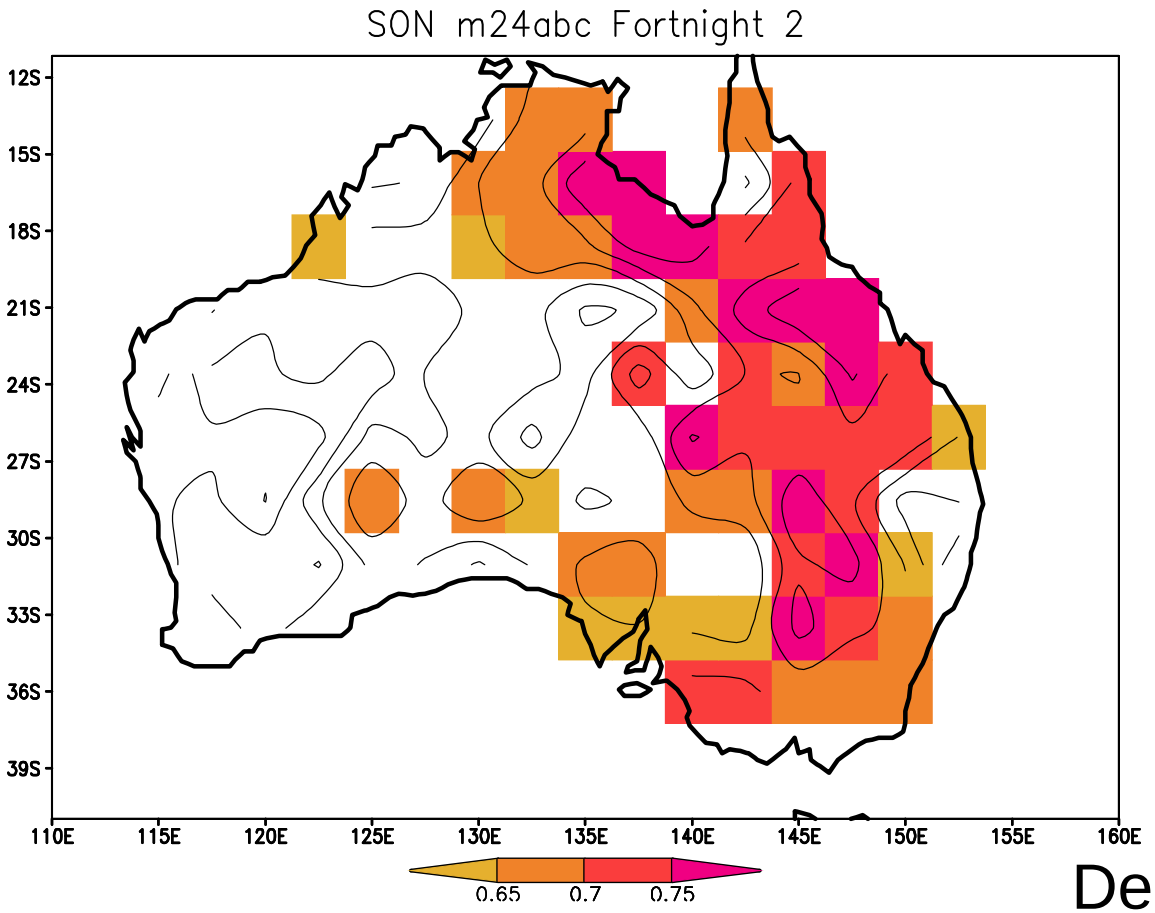
Persistence of day 5-18




Frederic Vitard and Laura Ferranti, ECMWF

# Two weeks forecast example: ½ month lead precip. fcsts

**ROC area: Precipitation anomalies in the upper tercile**
**Fortnight 2: Sep, Oct, Nov forecast start months. Hindcasts: 1980-2006**



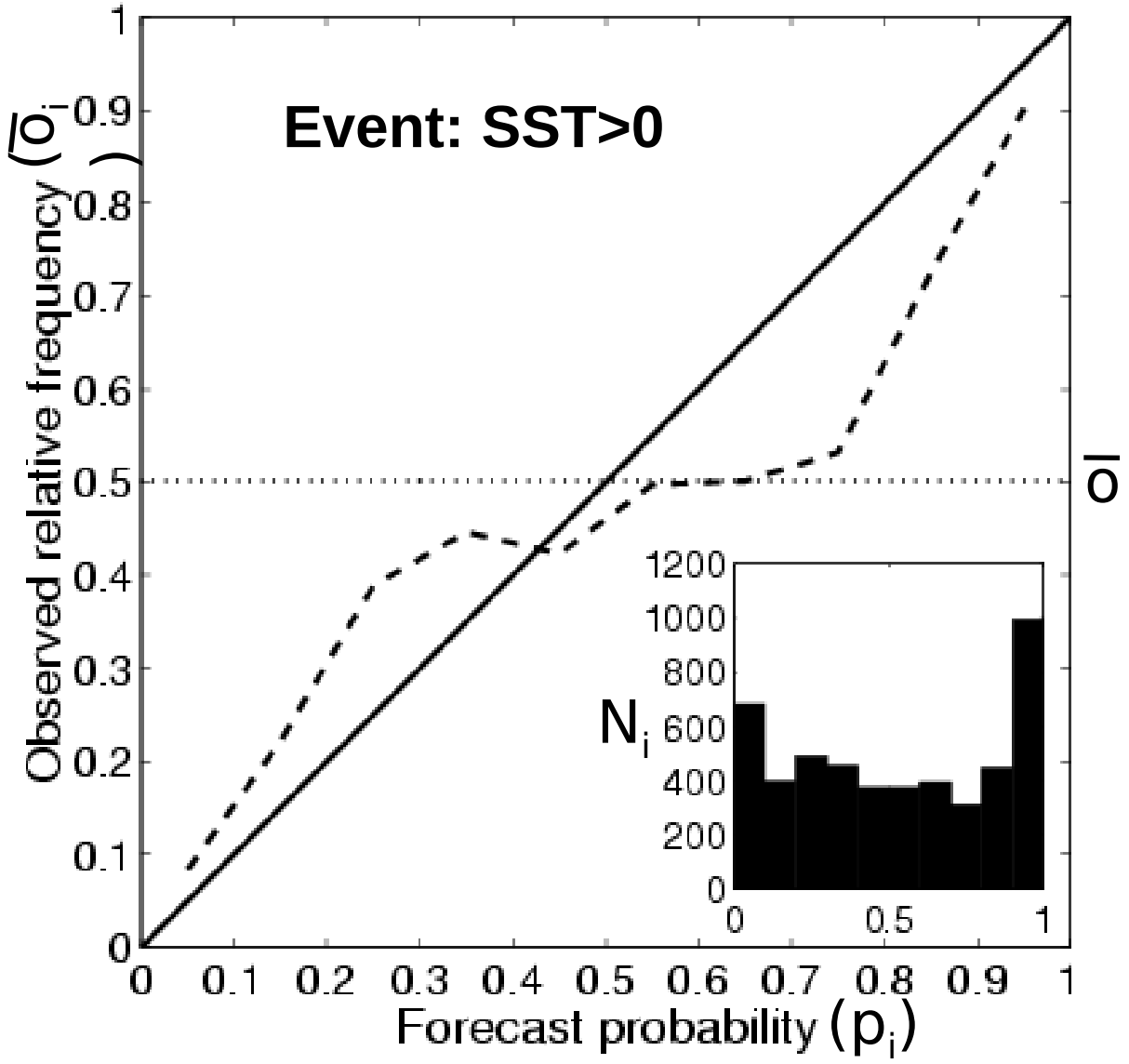

Debbie Hudson
BOM, Australia

# Reliability and resolution

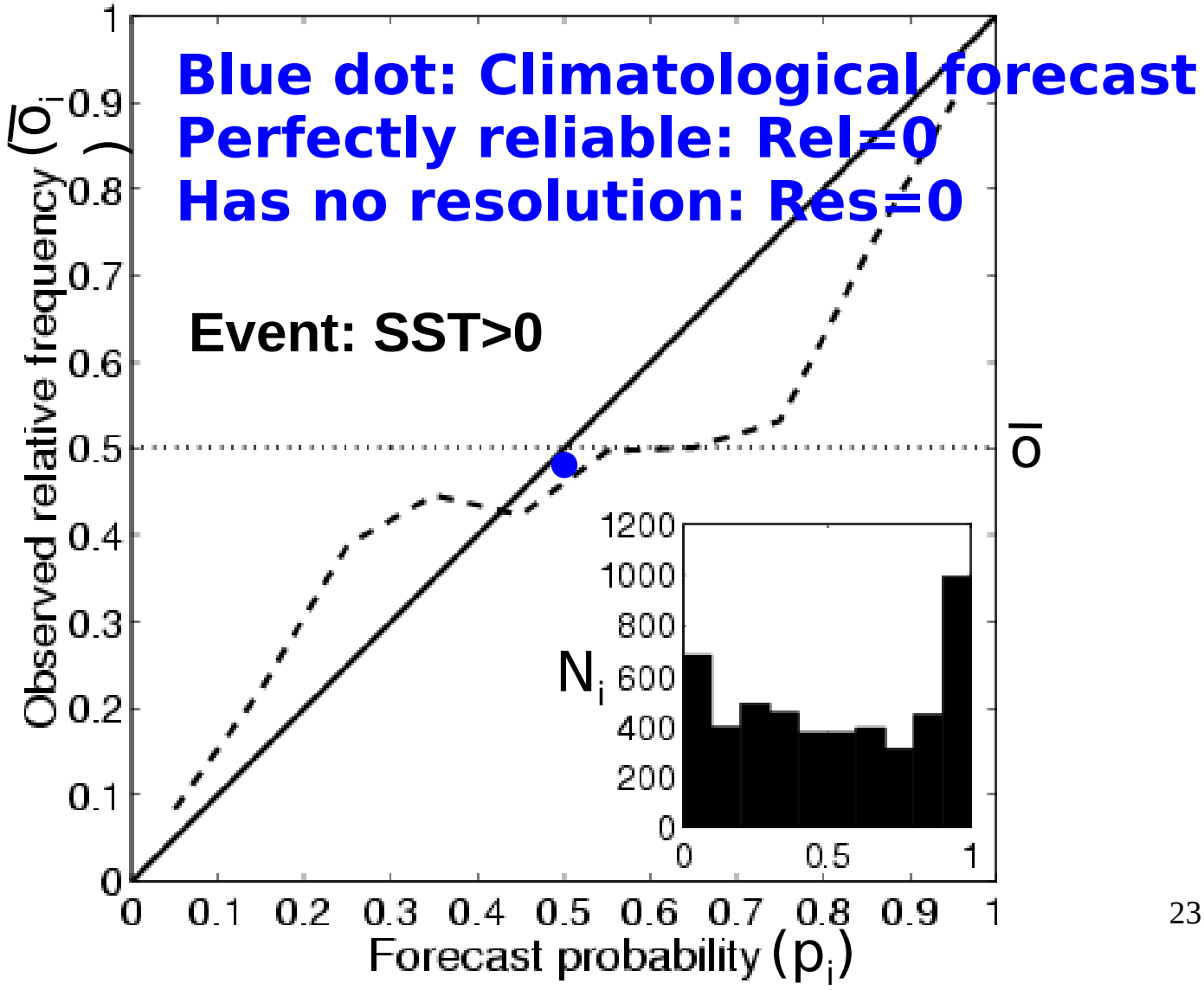
- Reliability: correspondence between forecast probabilities and observed relative frequency (e.g. an event must occur on 30% of the occasions that the 30% forecast probability was issued)

- Resolution: Conditioning of observed outcome on the forecasts
- Addresses the question: Does the frequency of occurrence of an event differs as the forecast probability changes?
- If the event occurs with the same relative frequency regardless of the forecast, the forecasts are said to have no resolution
- Forecasts with no resolution are useless because the outcome is the same regardless of what is forecast

# Reliability diagram



# Reliability diagram



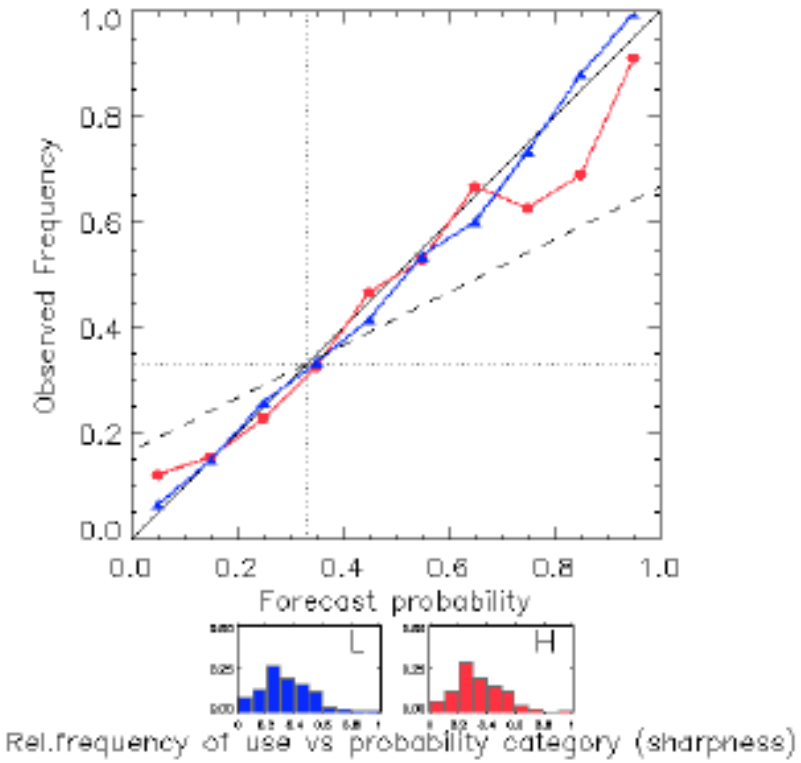
# Seasonal forecast example:

# 1-month lead MSLP fcsts for DJF

## GLOSEA5 Hindcast Probabilistic skill

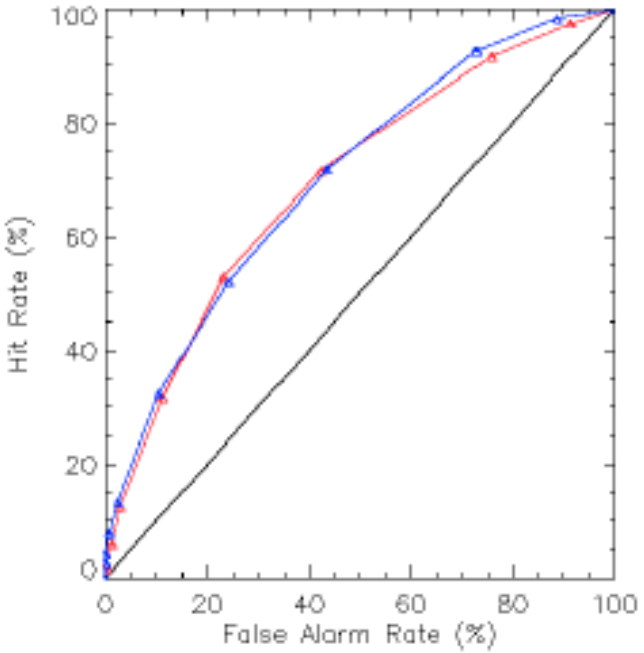
## MSLP in N. Atlantic in upper and lower tercile

**Reliability**




(a) Reliability diagram for mean sea level pressure in GloSea5 over the North Atlantic. The red line shows the upper tercile and the blue line is the lower tercile.

**ROC area**



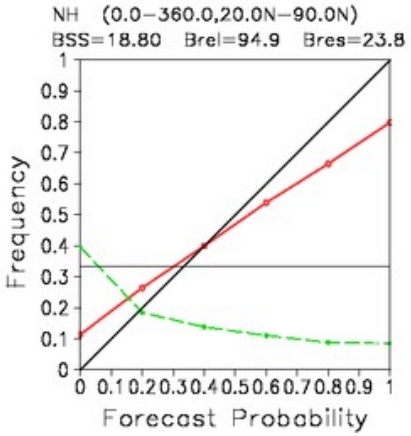

(b) Relative Operating Characteristics (ROC) diagram for the mean sea level pressure in GloSea5 over the North Atlantic. The red line shows the upper tercile and the blue line is the lower tercile.

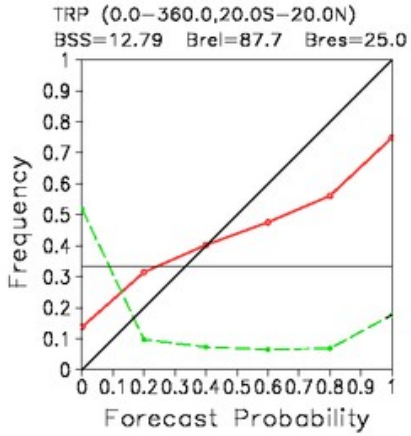
**Figure 6.** Statistical scores for the Northern Atlantic region.

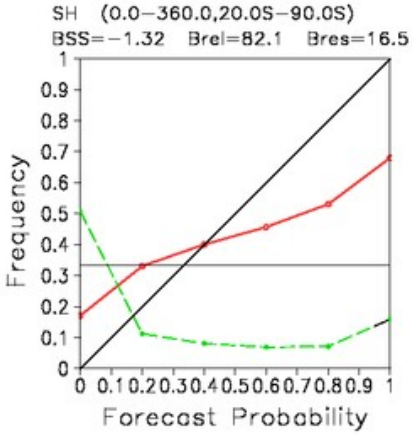
MacLachlan et al., QJRMS, 2015

# Monthly forecast example:
# 2-day lead 2mT fcsts for day 2-29 mean

< Reliability Diagram >
Event : T2m Anomaly Upper Tercile 2–29 day mean (V1403 vs JRA55)
BSS, Brel,Bres for 30 years (1981–2010) mem:5
Initial : DJF , Lead time : 2 day
Full(Red)=Reliability Dash(Green)=Forecast Frequency Brier Skill Scores x 100



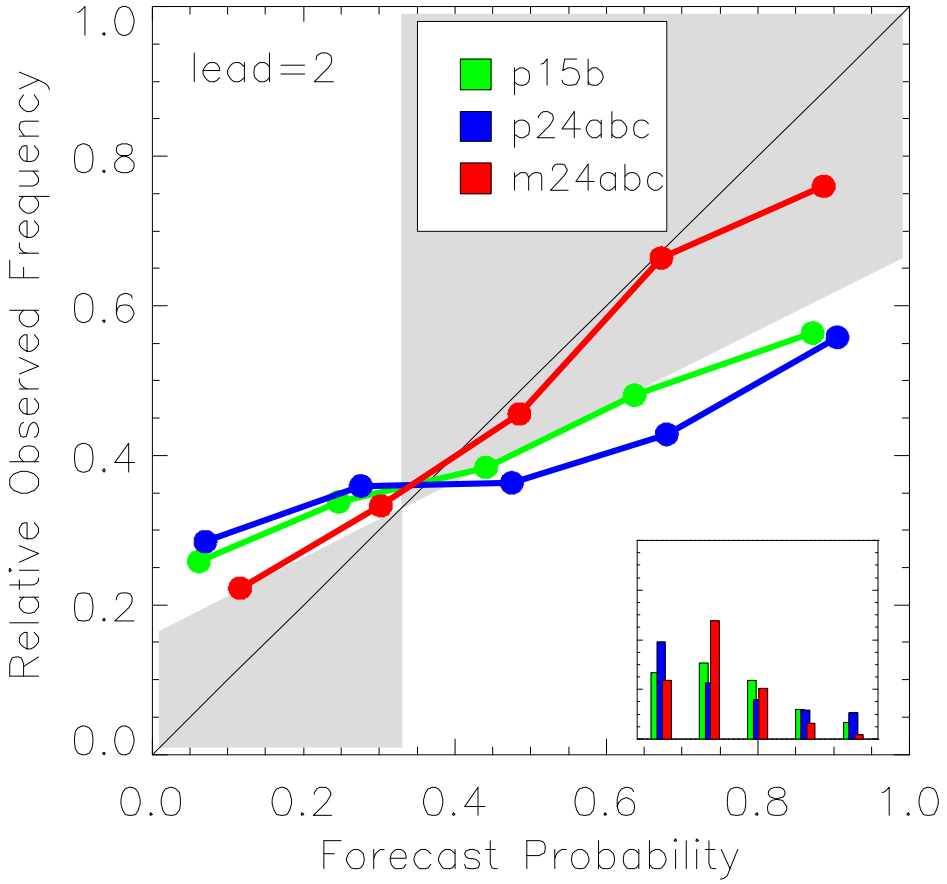




**Reliability Diagrams**
T2m (upper tercile)
Day 2-29 mean
I.C. : Dec.-Feb. 1981-2010
N.H., TROP, S.H.

Yuhei Takaya, JMA

# Two weeks forecast example: ½ month lead precip. fcsts

**Precipitation anomalies in the upper tercile**
**Fortnight 2: Sep, Oct, Nov forecast start months. Hindcasts: 1980-2006**



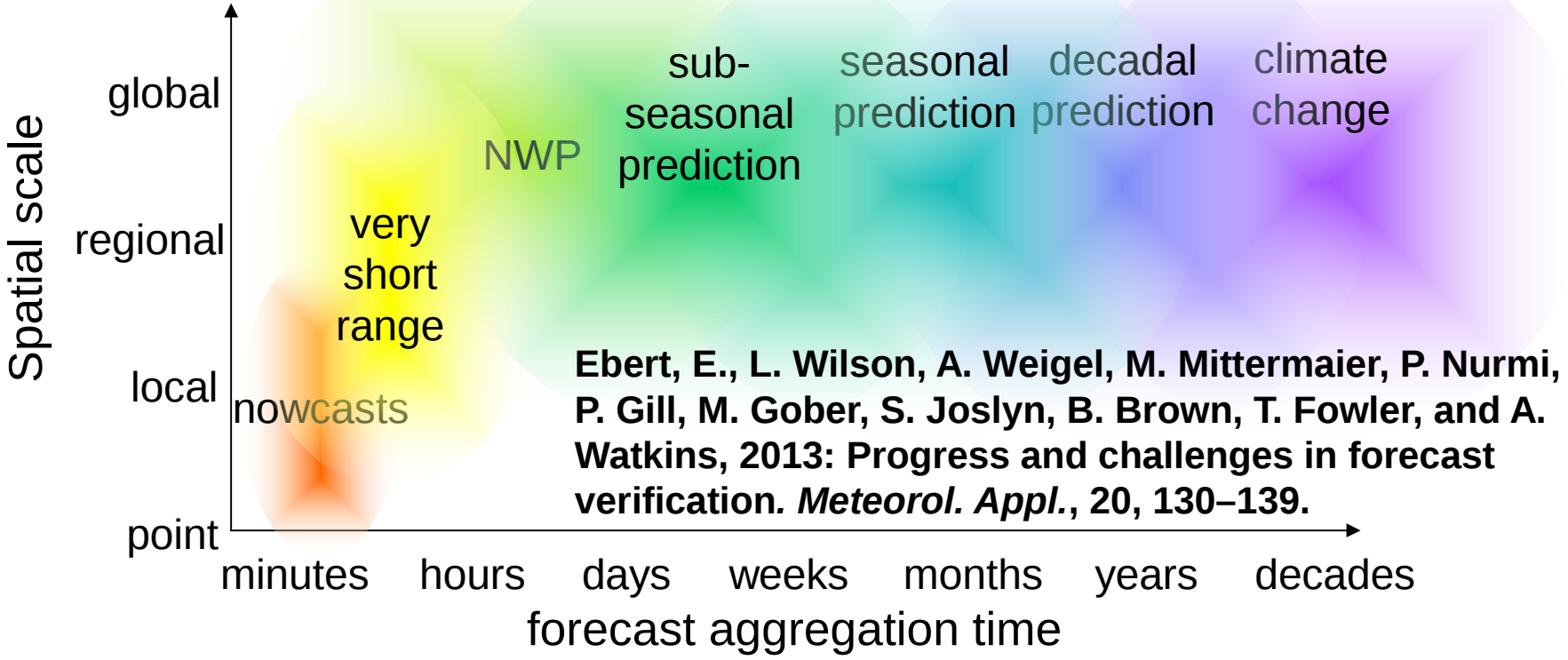

Debbie Hudson
BOM, Australia

# Seamless verification

Seamless forecasts - consistent across space/time scales

single modelling system or blended

probabilistic / ensemble




**Ebert, E., L. Wilson, A. Weigel, M. Mittermaier, P. Nurmi, P. Gill, M. Gober, S. Joslyn, B. Brown, T. Fowler, and A. Watkins, 2013: Progress and challenges in forecast verification. *Meteorol. Appl.*, 20, 130–139.**

# Final remarks

- Clear need for attributes-based verification for a complete forecast quality view

- Need for use more than a single score for more detailed forecast quality assessment

- S2S verification is naturally leaning towards the seamless consistency concept addressing the question of which scales and phenomena are predictable

- As S2S covers various forecast ranges (days, weeks and months) it naturally allows seamless verification developments

**Thank you all for your attention!**