Verification of seasonal forecasts

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

Acknowledgments: David Stephenson, Simon Mason, JMA/MRI and WMO LC-LRFMME

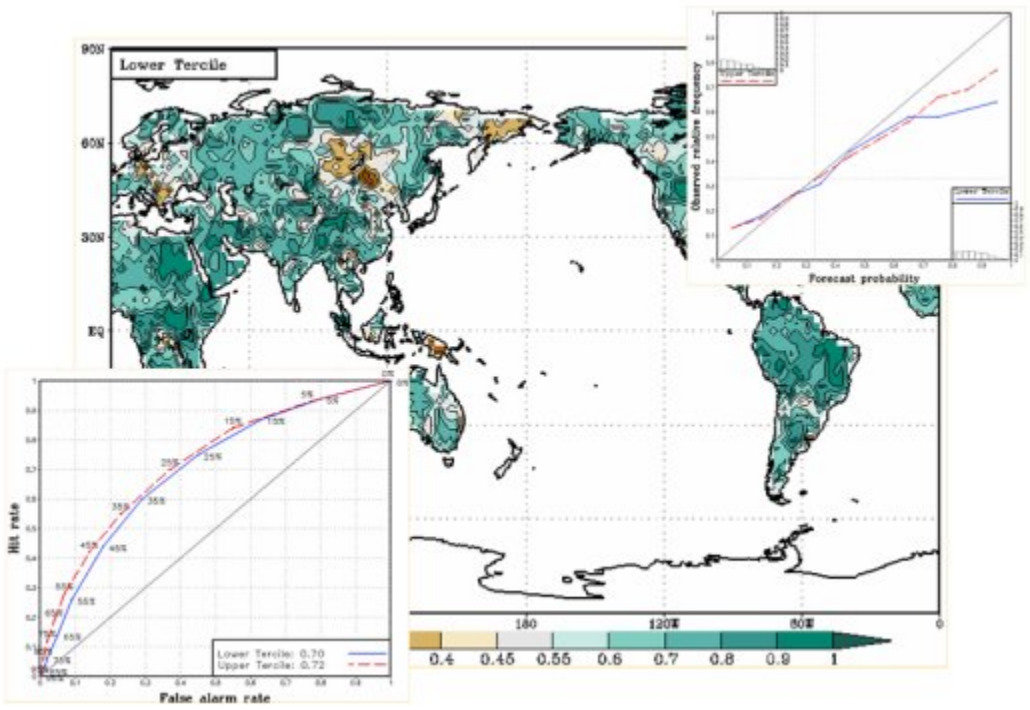
Plan of talk

1) Current WMO verification standards for seasonal forecast model outputs: GPCs-LRF
2) What is a good forecast?
3) Attributes based forecast quality assessment: definitions and examples of practice
4) WMO guidance on seasonal forecast verification: RCOFs and NMHS
5) Final remarks

Second WMO workshop on operational climate prediction
Barcelona, Spain, 30 May - 1 June 2018

WMO Standardized verification system for long-range forecasts (SVSLRF)

- Protocols for verification (forecast quality assessment) of WMO Global Producing Centres for Long-Range forecasts (GPCs-LRF) retrospective model outputs (hindcasts) are defined in the WMO Standardized Verification System for Long-Range Forecasts: Manual on GDPFS WMO No. 485
- Documentation and mathematical formulation of recommended scores available through the Lead Center for Long Range Forecast Verification (LC-SVSLRF)
- Useful to answer the question: How good were the retrospective dynamical model forecasts produced for a number of past years?



World Meteorological Organization
Lead Centre for the
Long Range Forecast Verification System

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

Quality: Measure of correspondence between forecasts and observations using mathematical relationships (deterministic and probabilistic scores)

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.

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

Accuracy

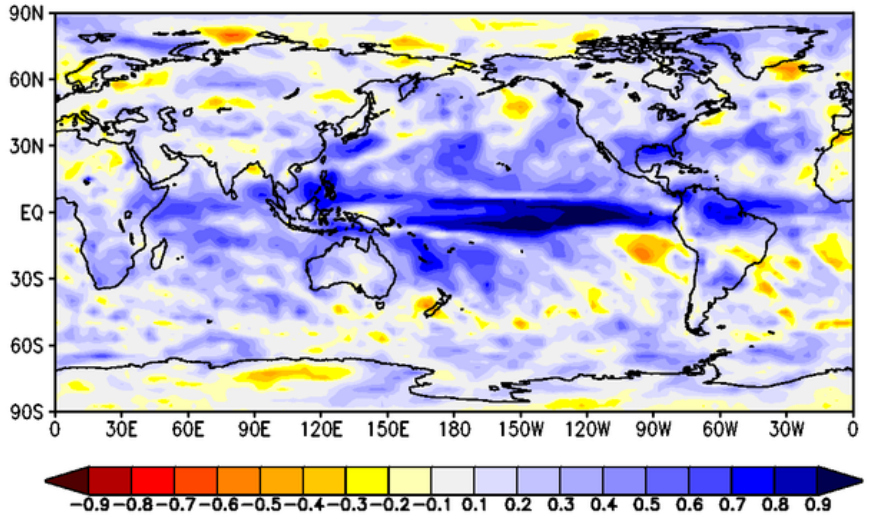
- Average distance between forecasts and observations
- Simplest measure is the **Mean Error (Bias)**

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

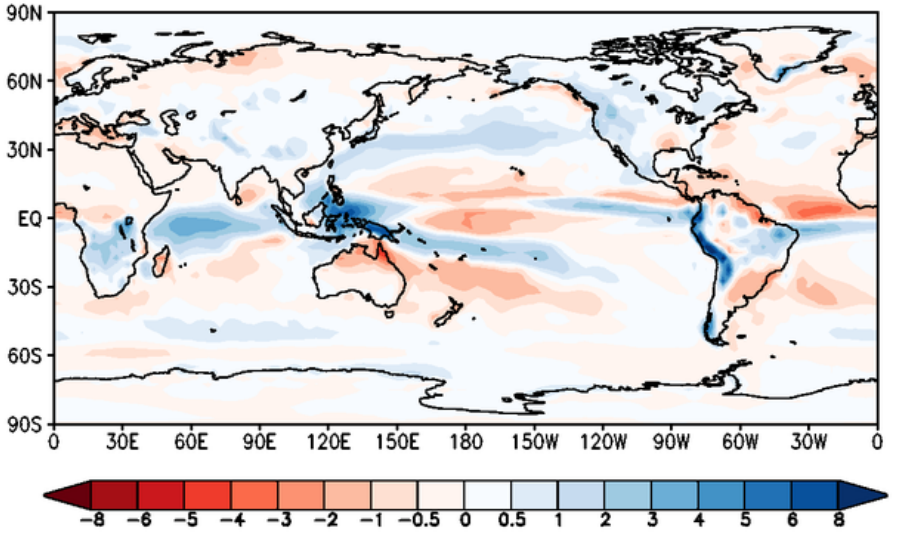
x: forecast y: observation n: number of (x,y) pairs

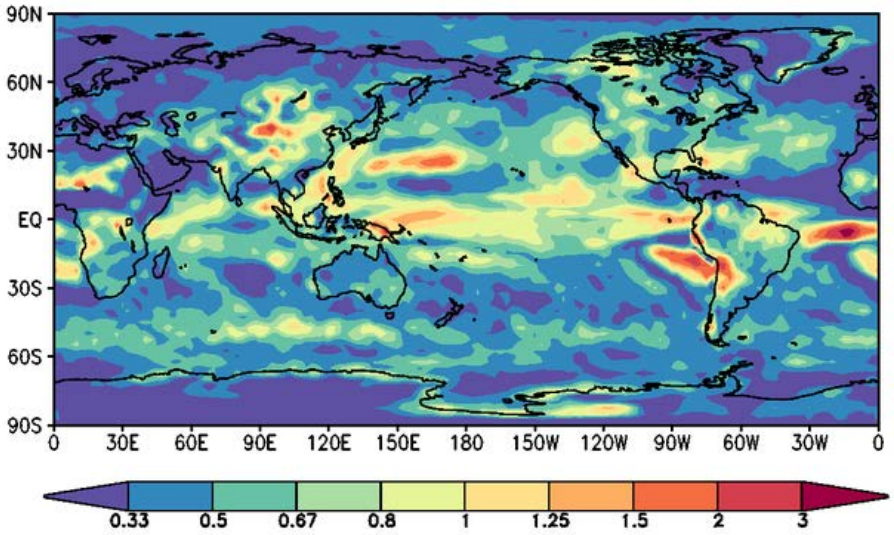
Example: JMA 1-month lead precip. fcsts for DJF

ACC (against GPCP v2.2)
I.C: Nov. Valid: DJF (1981-2010)



Bias (against GPCP v2.2)
I.C: Nov Valid: DJF (1981-2010)





St. dev ratio (F/O) (against GPCP v2.2)
I.C: Nov Valid: DJF (1981-2010)

Source: JMA/MRI

Forecast quality assessment

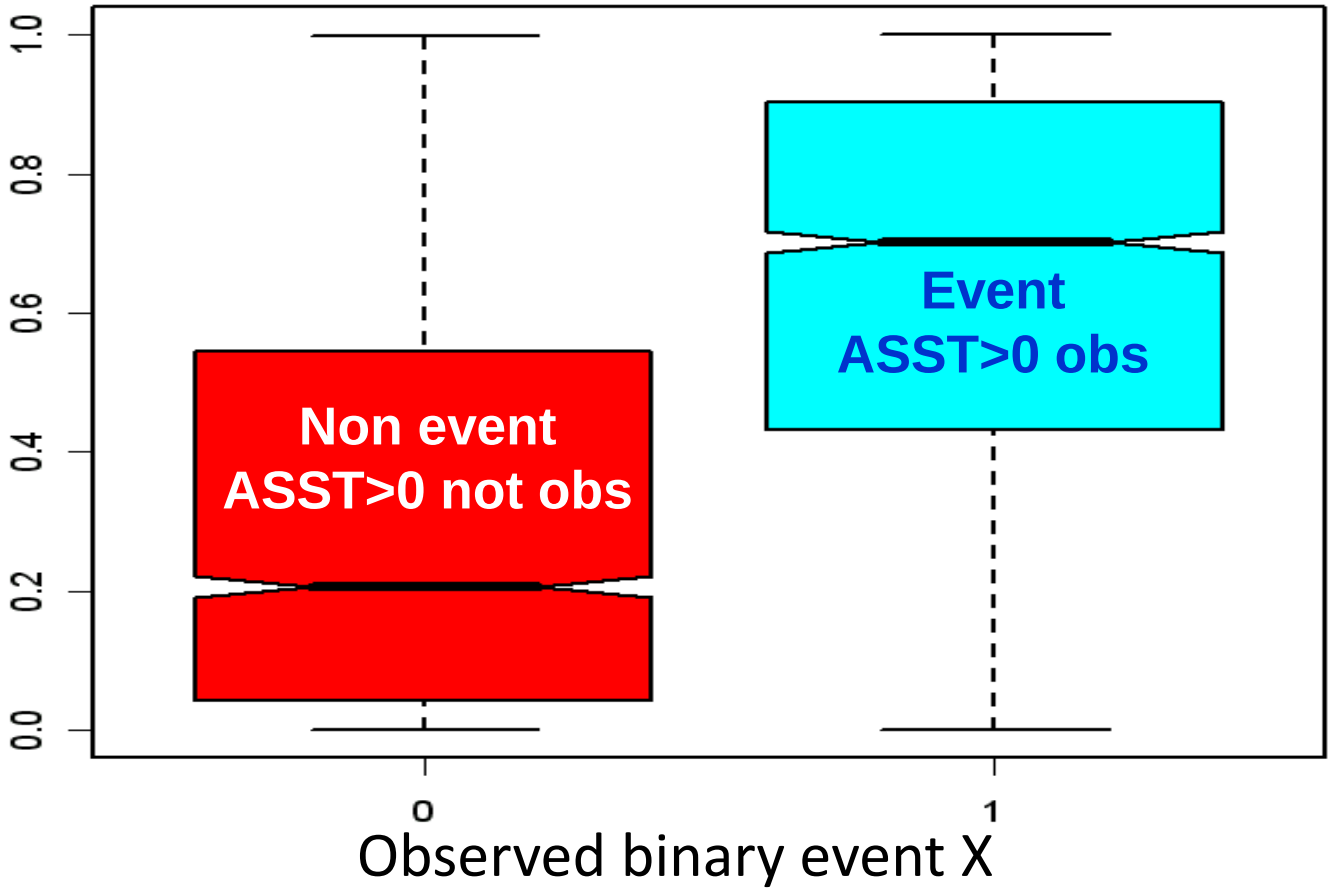
2. Attributes of probabilistic forecasts (derived from ensemble members)

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

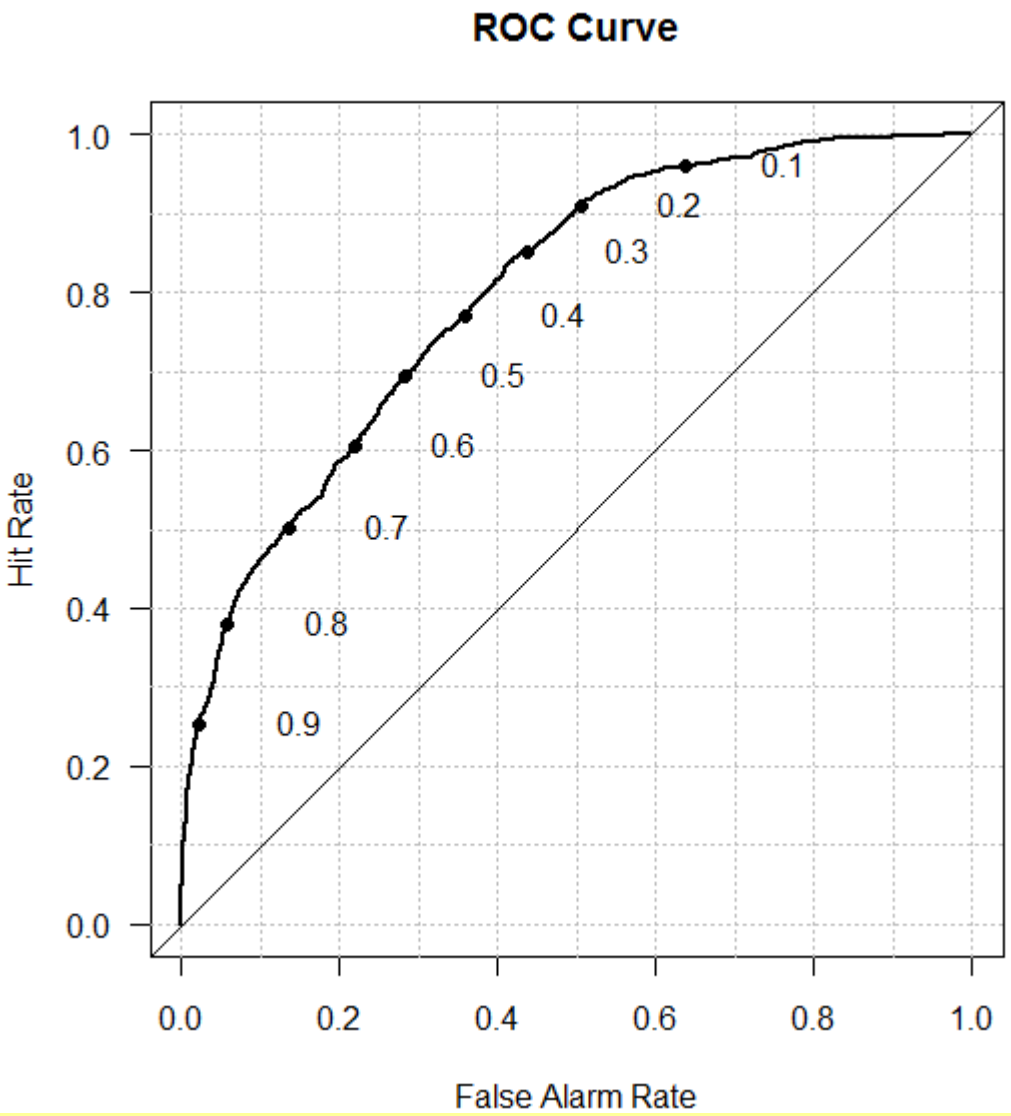
Prob. forecasts conditioned/stratified on observations

Forecast probability Pr(ASST>0) for a collection of grid points (Eq. Pacific)



→ Forecasts do differ given different outcomes

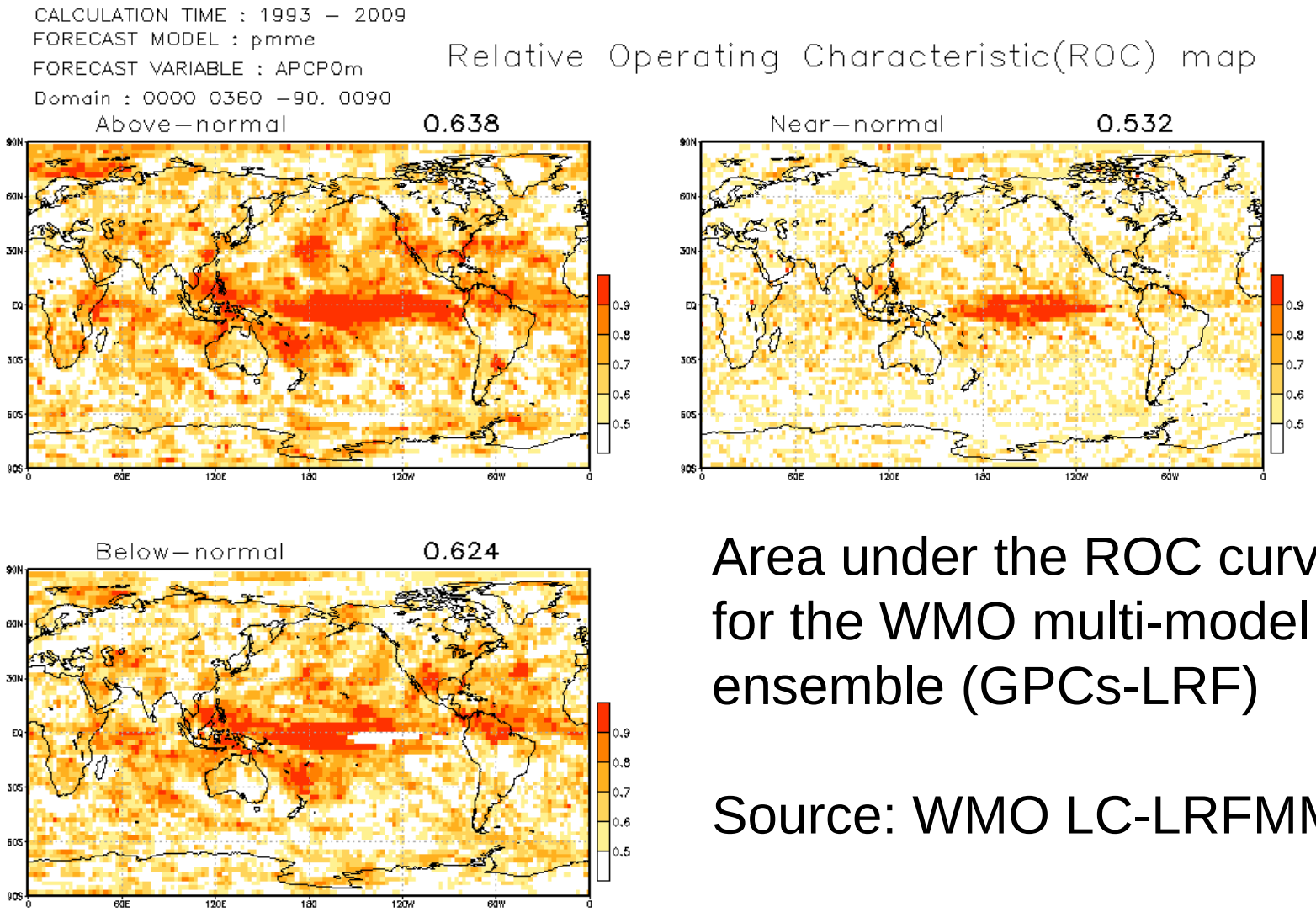
→ Forecast system has discrimination (distinguish event from non-event)



• The ROC curve is constructed by calc. the hit and false-alarm rates for decreasing prob. thresholds
• Area under ROC curve (A) is a measure of discrimination: A = 0.79 (prob. of successfully discriminating a warm (ASST>0) from a cold (ASST<0) event)
• As the events are binary (i.e. have two possible outcomes) the reference probability of successfully discriminating events from non-events is 50% (area under 45^{o} “no discrimination ”line)

LC-LRFMME centralized verification effort

1-month lead precip. fcsts for DJF



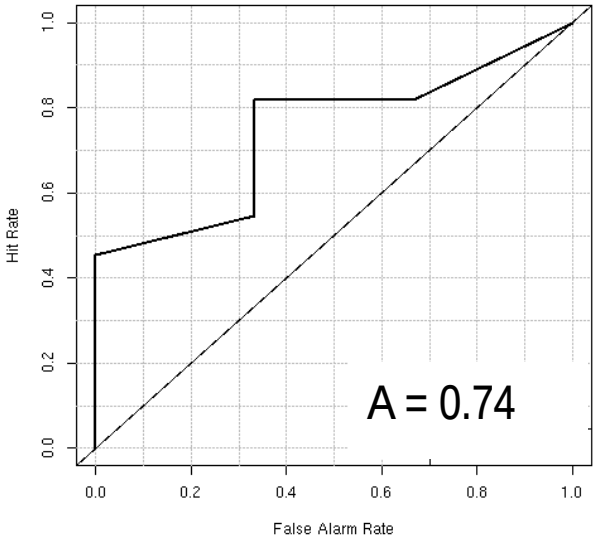
Area under the ROC curve (A) for the WMO multi-model ensemble (GPCs-LRF)

Source: WMO LC-LRFMME

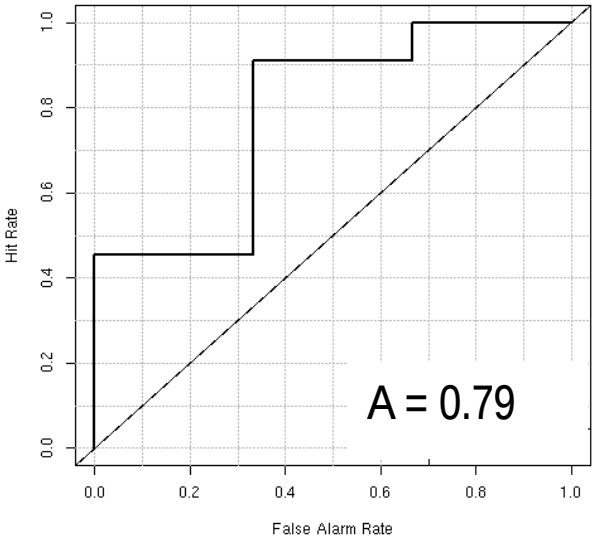
beijing/cptec/ecmwf/exeter/melbourne/montreal/moscow/offenbach/pretoria/seoul/tokyo/washington

Probabilistic rainy season onset forecast discrimination assessment for São Paulo, Brazil: 1996-2009

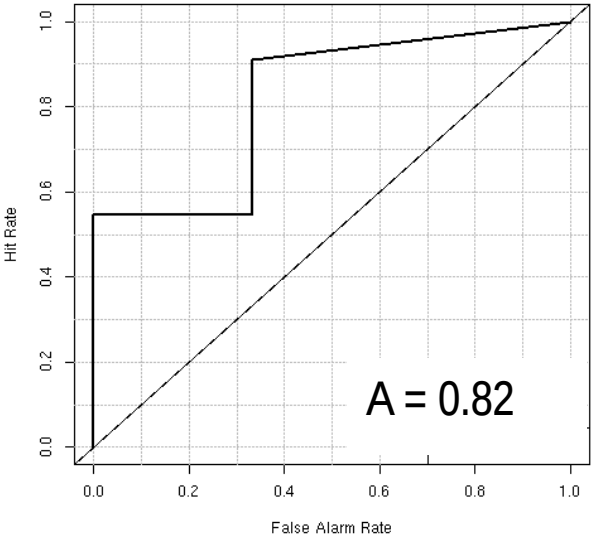
Empirical model:
Cox regression



Dynamical model:
GloSea 5



Combined:
Empirical and GloSea 5



• Use of daily precip. outputs from a seasonal forecast model (GloSea 5) to investigate the feasibility of issuing forecast probabilities that discriminate early from late onsets

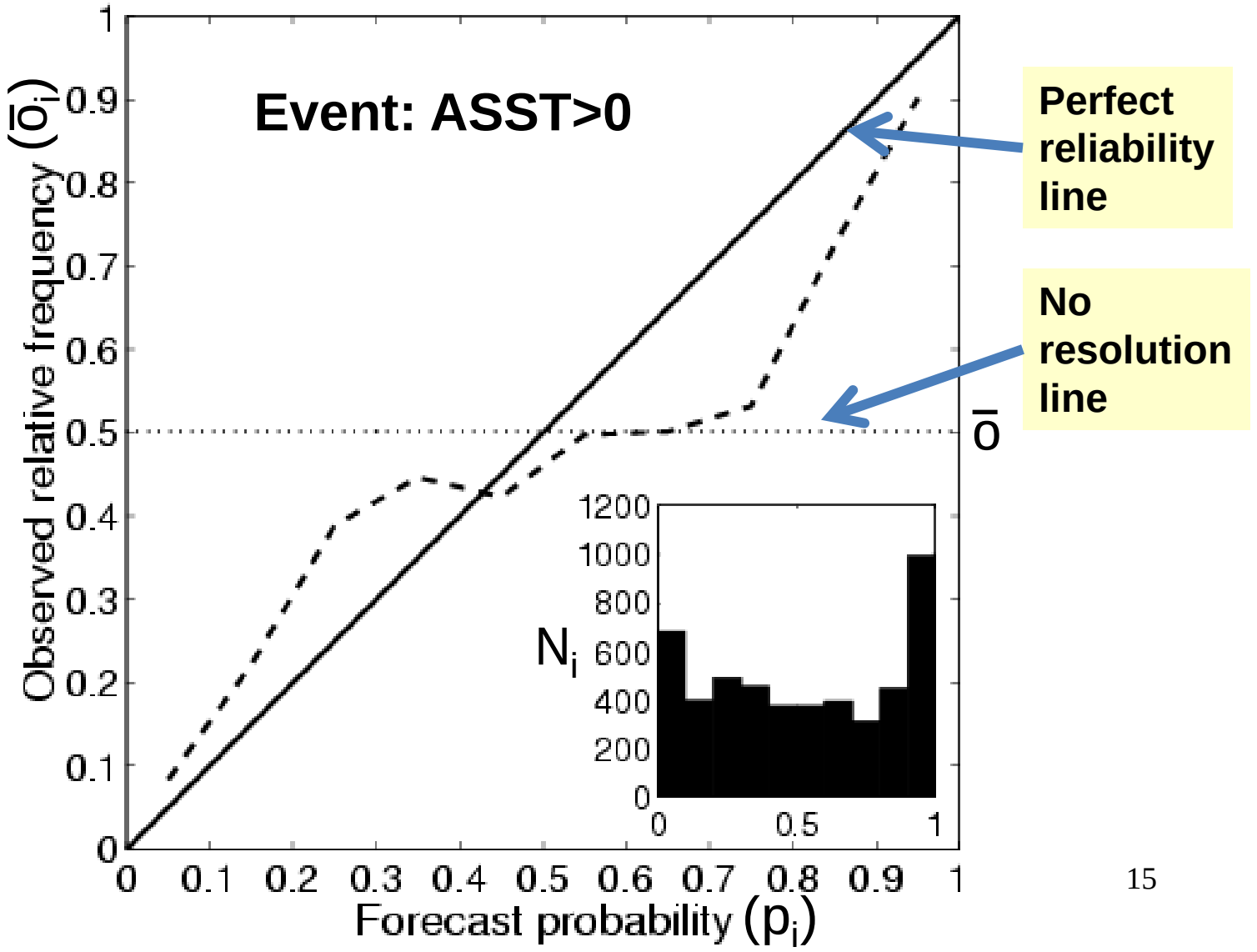
Coelho C.A.S., M.A.F. Firpo, A.H.N. Maia, C. MacLachlan, 2017: Exploring the feasibility of empirical, dynamical and combined probabilistic rainy season onset forecasts for São Paulo, Brazil. International Journal of Climatology.

Reliability and resolution

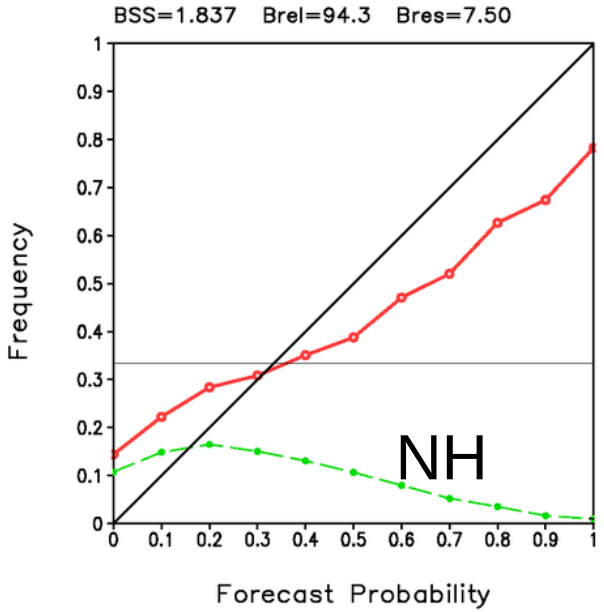
- Reliability: correspondence between forecast probabilities and observed relative frequency (e.g. an event must occur on 30% of the occasions that the 30% forecast probability was issued)

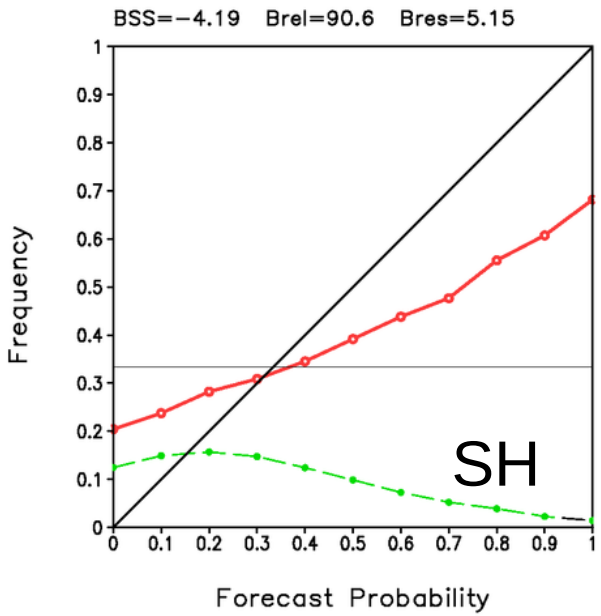
- Resolution: Conditioning of observed outcome on the forecasts
- Addresses the question: Does the frequency of occurrence of an event differs as the forecast probability changes?
- If the event occurs with the same relative frequency regardless of the forecast, the forecasts are said to have no resolution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

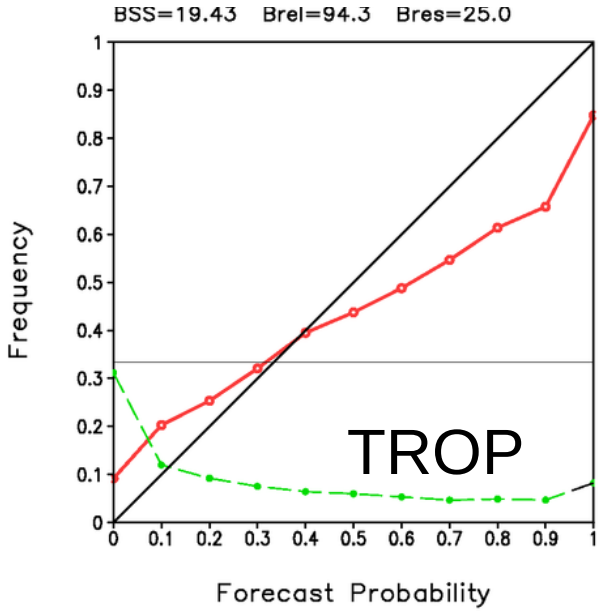
Reliability diagram



Example: JMA 1-month lead 2mT fcsts for DJF







Reliability Diagrams
T2m (upper tercile)
I.C. : Nov
Valid: DJF (1981-2010)
N.H., TROP, S.H.

Source: JMA/MRI

WMO guidance on verification of operational seasonal climate forecasts

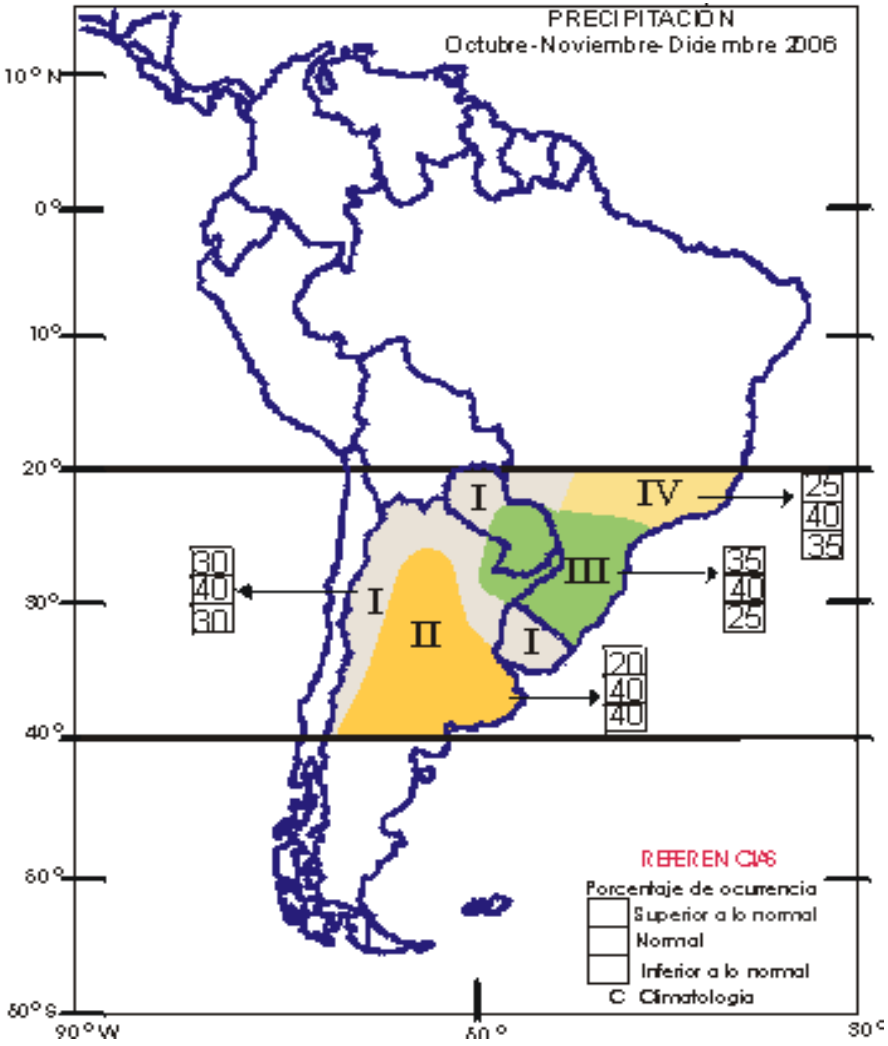
- Prepared by Simon Mason (IRI) under the auspices of the World Meteorological Organization (WMO) Commission for Climatology (CCl)
- Complementary to the WMO SVSLRF
- Recommendations for verification (largely focused on forecast quality assessment) of forecasts produced by RCOFs, NMHS and other forecasting centres (e.g. RCCs)
- Focused on probabilistic forecasts: model outputs, expert subjective assessment, or a combination of both
- Includes recommendations for verification of forecasts for one specific target period (e.g. last year RCOF forecast)
- Useful to answer questions such as: How good were the RCOF forecasts produced over a number of past years? How good was this RCOF forecast?

Verification of RCOF forecasts

One loud and clear result: hedging on normal.

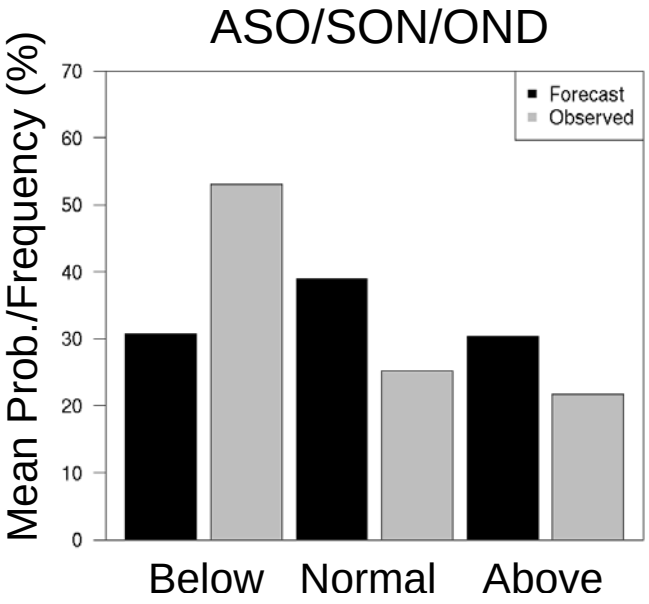
Do we genuinely think that normal is almost always the most likely category, or do we think it is the safest forecast?

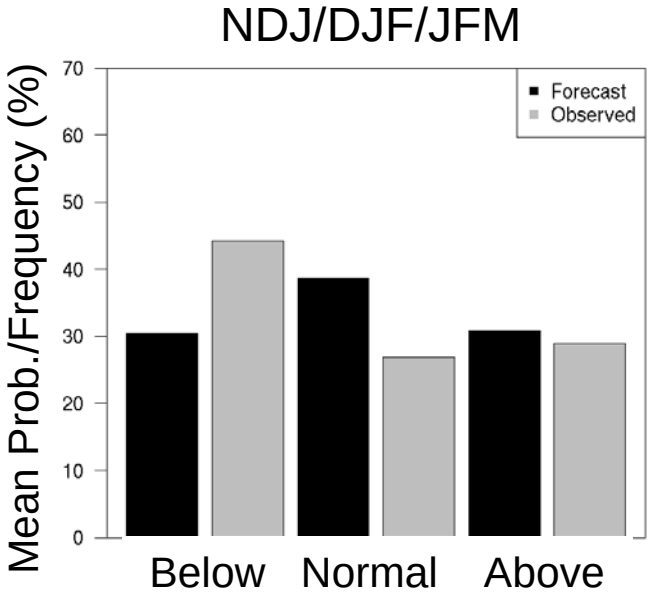
70 – 80% of all the African RCOF forecasts had highest probability on normal.

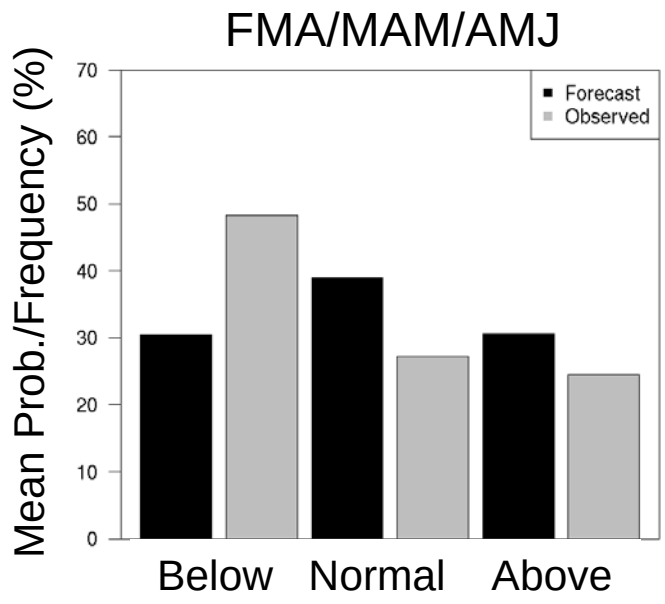


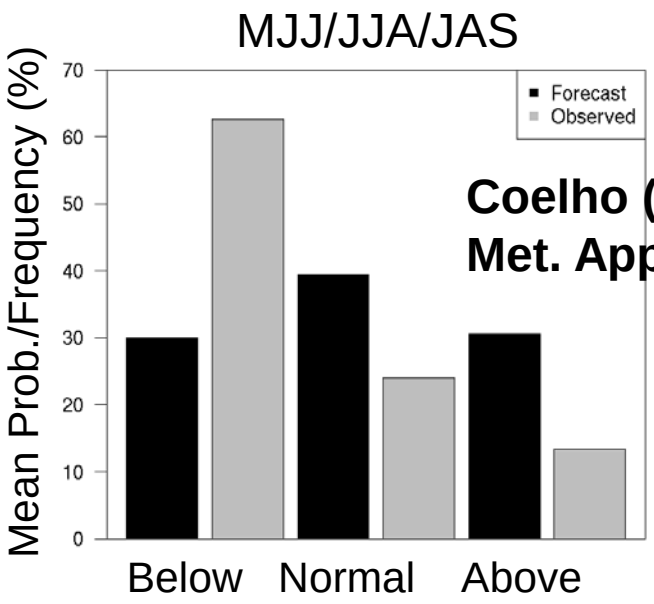
Courtesy: Simon Mason

Assessment of consensus forecast tercile probabilities for Brazil over 10 years









Coelho (2013)
Met. Applications

Final remarks

• Attributes-based verification is fundamental for a complete forecast quality view: need to use various scores for a detailed forecast quality assessment

• WMO standards for coordinated seasonal forecast verification already exist (SVSLRF): focused on model output retrospective forecasts (hindcasts)

• LC-LRFMME provides centralized verification (based on SVSLRF) for individual GPCs-LRF and WMO multi-model ensemble over a common hindcast period (currently 1993-2009)

• WMO/CCl Guidance on verification of operational seasonal forecasts produced by RCOFs, NMHSs and other forecasting centres (e.g. RCCs) already exist

• Scope for the assessment of historical seasonal forecasts produced by RCOFs over the last 20 years using the WMO/CCl guidance

• Scope for advancing seasonal climate prediction practice by better integrating model forecasts and verification information when issuing seasonal forecasts (RCOFs, NMHSs) through objective procedures

• Scope for developing WMO verification standards for emerging activities (sub-seasonal and annual to decadal predictions) and also for linking verification activities performed by research and operational communities

Thank you all for your attention!

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